

Министерство науки и высшего образования Российской Федерации  
 федеральное государственное автономное  
 образовательное учреждение высшего образования  
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Инженерная школа информационных технологий и робототехники  
 Направление подготовки 09.04.04 Программная инженерия  
 Отделение школы (НОЦ) Информационных технологий

### МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
<b>Алгоритмическое и программное обеспечение поиска аномалий в результатах медицинских исследований</b>

УДК 004.421:004.415.2:616-07

Студент

Группа	ФИО	Подпись	Дата
8ПМ9И	Попенко Анастасия Александровна		21.06.2021 г.

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент ОИТ ИШИТР	Аксёнов С.В.	к.т.н.		21.06.2021 г.

### КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент ОСГН ШБИП	Гончарова Н. А.	к.э.н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент ООД ШБИП	Антоневич О. А.	к.б.н.		

### ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
доцент ОИТ ИШИТР	Савельев А.О.	к.т.н.		

**ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ООП**  
по направлению 09.04.04 «Программная инженерия»

<b>Код компетенции</b>	<b>Наименование компетенции</b>
<b>Универсальные компетенции</b>	
УК(У)-1	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий
УК(У)-2	Способен управлять проектом на всех этапах его жизненного цикла
УК(У)-3	Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели
УК(У)-4	Способен применять современные коммуникативные технологии, в том числе на иностранном (-ых) языке (-ах), для академического и профессионального взаимодействия
УК(У)-5	Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия
УК(У)-6	Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки
<b>Общепрофессиональные компетенции</b>	
ОПК(У)-1	Способен самостоятельно приобретать, развивать и применять математические, естественно-научные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте
ОПК(У)-2	Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач
ОПК(У)-3	Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями
ОПК(У)-4	Способен применять на практике новые научные принципы и методы исследований

ОПК(У)-5	Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем
ОПК(У)-6	Способен самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе в новых областях знаний, непосредственно не связанных со сферой деятельности
ОПК(У)-7	Способен применять при решении профессиональных задач методы и средства получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе, в глобальных компьютерных сетях
ОПК(У)-8	Способен осуществлять эффективное управление разработкой программных средств и проектов
<b>Профессиональные компетенции</b>	
ПК(У)-1	Способен к созданию вариантов архитектуры программного средства
ПК(У)-2	Способен разрабатывать и администрировать системы управления базам данных
ПК(У)-3	Способен управлять процессами и проектами по созданию (модификации) информационных ресурсов
ПК(У)-4	Способен проектировать и организовывать учебный процесс по образовательным программам с использованием современных образовательных технологий
ПК(У)-5	Способен осуществлять руководство разработкой комплексных проектов на всех стадиях и этапах выполнения работ

Министерство науки и высшего образования Российской Федерации  
 федеральное государственное автономное  
 образовательное учреждение высшего образования  
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Инженерная школа информационных технологий и робототехники  
 Направление подготовки (специальность) 09.04.04 Программная инженерия  
 Отделение школы (НОЦ) Информационных технологий

УТВЕРЖДАЮ:  
 Руководитель ООП  
 \_\_\_\_\_ Савельев А.О.  
 (подпись) (дата) (Ф.И.О.)

### ЗАДАНИЕ на выполнение выпускной квалификационной работы

В форме:

Магистерской диссертации
--------------------------

(бакалаврской работы, дипломного проекта/работы, магистерской диссертации)

Студенту:

Группа	ФИО
8ПМ9И	Попенко Анастасия Александровна

Тема работы:

Алгоритмическое и программное обеспечение поиска аномалий в результатах медицинских исследований	
Утверждена приказом директора (дата, номер)	№ 40-5/с от 09.02.2021

Срок сдачи студентом выполненной работы:	21.06.2021
--	------------

#### ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

<b>Исходные данные к работе</b>  <i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).</i>	1. Медицинские исследования, формируемые на основе осмотров пациентов лечащим врачом. 2. Для обучения модели доступны 1 500 тыс. данных медицинских исследований предоставленные СибГМУ и скачаны с публичной платформы Kaggle. 3. Требование создать алгоритм для анализа и визуализации аномалий.
--	---

<p><b>Перечень подлежащих исследованию, проектированию и разработке вопросов</b></p> <p><i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i></p>	<p>1. Обзор алгоритмов. 2. Исследование предметной области. 3. Проектирование. 4. Тестирование. 5. Работа над разделом по финансовому менеджменту, ресурсоэффективности и ресурсосбережения. 6. Работа над разделом по социальной ответственности. 7. Работа над разделом на английском языке.</p>
<p><b>Перечень графического материала</b></p> <p><i>(с точным указанием обязательных чертежей)</i></p>	<p>Презентация в формате *.pptx</p>
<p><b>Консультанты по разделам выпускной квалификационной работы</b></p> <p><i>(с указанием разделов)</i></p>	
<p><b>Раздел</b></p>	<p><b>Консультант</b></p>
<p>Основная часть</p>	<p>Доцент ОИТ ИШИТР, к.т.н., доцент Аксёнов С.В.</p>
<p>Финансовый менеджмент, ресурсоэффективность и ресурсосбережение</p>	<p>Доцент ОСГН ШБИП, к.э.н., доцент Гончарова Н.А.</p>
<p>Социальная ответственность</p>	<p>Доцент ООД ШБИП, к.б.н., доцент Антоневиц О. А.</p>
<p>Английский язык</p>	<p>Доцент ОИЯ ШБИП, к.п.н., доцент Сидоренко Т.В.</p>
<p><b>Названия разделов, которые должны быть написаны на русском и иностранном языках:</b></p>	
<p>Раздел 3 Object and research methods</p>	

<p><b>Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику</b></p>	<p>01.03.2021</p>
--	-------------------

**Задание выдал руководитель:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент ОИТ ИШИТР	Аксёнов С.В.	к.т.н., доцент		01.03.2021

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ПМ9И	Попенко Анастасия Александровна		01.03.2021

Министерство науки и высшего образования Российской Федерации  
 федеральное государственное автономное  
 образовательное учреждение высшего образования  
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Инженерная школа информационных технологий и робототехники  
 Направление подготовки (специальность) 09.04.04 Программная инженерия  
 Уровень образования магистратура  
 Отделение школы (НОЦ) Информационных технологий  
 Период выполнения весенний семестр 2020 /2021 учебного года

Форма представления работы:

Магистерская диссертация
--------------------------

(бакалаврская работа, дипломный проект/работа, магистерская диссертация)

### КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:	21.06.2021
--	------------

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
01.06.2021	Основная часть	70
01.06.2021	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	10
01.06.2021	Социальная ответственность	10
01.06.2021	Английский язык	10

#### СОСТАВИЛ:

##### Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент ОИТ ИШИТР	Аксёнов С.В.	к.т.н.		

#### СОГЛАСОВАНО:

##### Руководитель ООП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент ОИТ ИШИТР	Савельев А. О.	к.т.н.		

# ЗАДАНИЕ ДЛЯ РАЗДЕЛА «ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ»

Студенту:

Группа	ФИО
8ПМ9И	Попенко Анастасия Александровна

Школа	ИШИТР	Отделение школы (НОЦ)	ОИТ
Уровень образования	Магистратура	Направление/специальность	09.04.04 «Программная инженерия»

## Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	1. Оклад руководителя – 33 664 руб. 2. Оклад инженера – 15 470 руб. 3. Стоимость материальных ресурсов определялась согласно прейскурантам компаний.
2. Нормы и нормативы расходования ресурсов	-
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	Коэффициент отчислений во внебюджетные фонды – 30 %.

## Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Расчет инновационного потенциала НТИ	Определение цели научно-исследовательского проекта, описание потенциальных потребителей проекта и результатов его внедрения.
2. Расчет сметы затрат на выполнение проекта	1. Расчет материальных затрат. 2. Расчет основной и дополнительной заработной платы. 3. Расчет отчислений во внебюджетные фонды. 4. Расчет бюджета проекта.

## Перечень графического материала (с точным указанием обязательных чертежей):

1. «Портрет» потребителя результатов НТИ
2. Сегментирование рынка
3. Оценка конкурентоспособности технических решений
4. График проведения и бюджет НТИ
5. Оценка ресурсной, финансовой и экономической эффективности НТИ
6. Потенциальные риски

Дата выдачи задания для раздела по линейному графику	01.03.2021
--	------------

## Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент ОСГН ШБИП	Гончарова Наталья Александровна	к.э.н.		

## Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ПМ9И	Попенко Анастасия Александровна		

## ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
8ПМ9И	Попенко Анастасия Александровна

<b>Школа</b>	Инженерная школа информационных технологий и робототехники	<b>Отделение (НОЦ)</b>	Отделение информационных технологий
<b>Уровень образования</b>	Магистратура	<b>Направление/специальность</b>	09.04.04 «Программная инженерия»

Тема ВКР:

Алгоритмическое и программное обеспечение поиска аномалий в результатах медицинских исследований	
<b>Исходные данные к разделу «Социальная ответственность»:</b>	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Применение ряда методов выявления аномальных данных набора результатов медицинских анализов. Работа ведется в условиях учебной аудитории с использованием персонального компьютера.
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
<b>1. Правовые и организационные вопросы обеспечения безопасности:</b> <ul style="list-style-type: none"> <li>– специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства;</li> <li>– организационные мероприятия при компоновке рабочей зоны.</li> </ul>	Трудовой Кодекс Российской Федерации от 30.12.2001 N 197-ФЗ (ред. от 30.04.2021) Федеральный закон "О специальной оценке условий труда" от 28.12.2013 N 426-ФЗ "ГОСТ 12.2.032-78. Система стандартов безопасности труда. Рабочее место при выполнении работ сидя. Общие эргономические требования"
<b>2. Производственная безопасность:</b> 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия	Вредные факторы: –Отклонение показателей микроклимата; –Превышение уровня шума; –Отсутствие или недостаток естественного света; –Повышенная напряженность магнитного поля; –Психофизиологический фактор; –Перенапряжение зрительного анализатора. Опасный фактор: –Электромагнитные излучения.
<b>3. Экологическая безопасность:</b>	Анализ воздействия на окружающую природную среду: утилизация люминесцентных ламп, компьютеров и другой оргтехники.
<b>4. Безопасность в чрезвычайных ситуациях:</b>	Удар молнии, взрыв, наводнение. Наиболее типичная ЧС – пожар.



Дата выдачи задания для раздела по линейному графику	01.03.2021
--	------------

**Задание выдал консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент ООД ШБИП	Антоневич Ольга Алексеевна	к.б.н.		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ПМ9И	Попенко Анастасия Александровна		

## РЕФЕРАТ

Выпускная квалификационная работа 85 с., 21 рис., 14 табл., 33 источника, 3 прил.

Ключевые слова: медицинские анализы, алгоритм, ансамблевые методы, обнаружение аномалий, python, машинное обучение.

Объектом исследования является (ются) медицинская документация, формируемая на основе осмотров пациентов лечащим врачом.

Цель работы – разработка алгоритмического и программного обеспечения анализа и визуализации результатов медицинских осмотров

В ходе исследования был проведен аналитический обзор по литературным источникам с целью уточнения достижений мировой технологической науки в рассматриваемой области; анализ и визуализация результатов медицинских осмотров; определение аномалий в результатах медицинских исследований.

В результате исследования проведены анализ и визуализация результатов медицинских осмотров; разработан алгоритм определения аномалий в результатах медицинских исследований.

Область применения: здравоохранение

Экономическая эффективность/значимость работы - применение алгоритма позволит автоматически определять аномалии в медицинских данных, что сократит время работы специалиста, затрачиваемое на данную процедуру.

В будущем планируется работа по внедрению алгоритма в медицинские системы здравоохранения.

## **ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ, СОКРАЩЕНИЯ**

Kaggle - система для проведения конкурсов по исследованию данных, а также социальная сеть специалистов по обработке данных и машинному обучению [7].

DM - DataMining

ML - Machine Learning

SL - Supervised learning

UL - Unsupervised learning

SSL - Semi-Supervised learning

RL - Reinforcement learning

DL - Deep learning

IF - IsolationForests

SVM - One Class SVM

EE - Elliptic Envelope

## **ОГЛАВЛЕНИЕ**

<b>РЕФЕРАТ .....</b>	<b>10</b>
<b>ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ, СОКРАЩЕНИЯ.....</b>	<b>11</b>
<b>ВВЕДЕНИЕ.....</b>	<b>15</b>
<b>ГЛАВА I. ОСНОВНАЯ ЧАСТЬ.....</b>	<b>17</b>
<b>Обзор литературы .....</b>	<b>17</b>
<b>1. Объект и методы исследования.....</b>	<b>20</b>
1.1 Методы машинного обучения.....	20
1.2 Обучение с учителем (Supervised learning).....	20
1.3 Обучение без учителя (Unsupervised learning) .....	20
1.4 Обучение с частичным привлечением учителя (Semi-Supervised learning) 21	
1.5 Обучение с подкреплением (Reinforcement learning) .....	21
1.6 Глубокое обучение (Deep learning) .....	22
1.7 Ансамблевые методы (Ensemble Methods) .....	22
<b>2. Расчёты и аналитика.....</b>	<b>24</b>
2.1 Выбор программного обеспечения .....	24
2.1.1 Используемые библиотеки «Python».....	24
2.2 Загрузка и предварительный анализ данных .....	25
2.3 Разделение набора данных на тестовую и тренировочную выборки ...	25
<b>3. Построение классификатора.....</b>	<b>26</b>
3.1 Выбор алгоритма классификации .....	26
3.1.1 Изолирующий лес (Isolation Forest).....	26
3.1.2 Одноклассовый метод опорных векторов (One Class SVM).....	27
3.1.3 Алгоритм эллиптической огибающей (Elliptic Envelope) .....	28
3.1.4 Ансамбли алгоритмов.....	29
<b>4. Результаты проведенного исследования .....</b>	<b>34</b>
<b>Заключение по разделу .....</b>	<b>40</b>
<b>ГЛАВА II. ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ .....</b>	<b>41</b>

Введение.....	41
1. Организация и планирование работ .....	41
2. Продолжительность этапов работ .....	42
3. Расчет сметы затрат на выполнение проекта.....	45
4. Расчет заработной платы .....	45
5. Расчет затрат на социальный налог .....	46
6. Расчет затрат на электроэнергию.....	47
6.1. Расчет амортизационных расходов .....	49
6.2. Расчет прочих расходов .....	49
7. Расчет общей себестоимости разработки .....	50
7.1. Расчет прибыли .....	50
7.2. Расчет НДС.....	50
7.3. Цена разработки НИР .....	51
8. Оценка экономической эффективности проекта .....	51
Заключение по разделу .....	51
<b>ГЛАВА III. СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ.....</b>	<b>52</b>
Введение .....	52
1. Правовые и организационные вопросы обеспечения безопасности.....	52
1.1. Специальные правовые нормы трудового законодательства.....	52
1.2. Эргономические требования к рабочему месту оператора ПЭВМ.....	53
2. Производственная безопасность .....	54
2.1. Анализ вредных и опасных факторов, которые может создать объект исследования.....	54
2.2. Обоснование мероприятий по снижению воздействия вредных и опасных факторов.....	55
2.2.1. Отклонение показателей микроклимата.....	55
2.3. Превышение уровня шума.....	57
2.4. Отсутствие или недостаток естественного и искусственного освещения.....	58
2.5. Повышенная напряженность магнитного поля .....	63

2.6. Экологическая безопасность .....	63
3.1. Анализ вероятных ЧС, которые может инициировать объект исследований.....	65
3.2. Меры по предотвращению и ликвидации ЧС и их последствий .....	66
Заключение по разделу .....	67
ЗАКЛЮЧЕНИЕ .....	68
СПИСОК ПУБЛИКАЦИЙ И НАУЧНЫХ ДОСТИЖЕНИЙ .....	69
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....	70
ПРИЛОЖЕНИЕ А .....	74
Object and research methods .....	75
1.1. Machine learning techniques .....	75
1.1.1. Supervised Learning .....	76
1.1.2. Unsupervised Learning .....	76
1.1.3. Reinforcement Learning .....	77
1.1.4. Deep Learning .....	77
1.1.5. Reinforcement Learning .....	78
1.1.6. Ensemble Methods .....	78
ПРИЛОЖЕНИЕ Б .....	80
ПРИЛОЖЕНИЕ В .....	84

## ВВЕДЕНИЕ

С момента появления информационных технологий во многих сферах деятельности были внедрены изменения, которые улучшают, оптимизируют и автоматизируют определённые процессы. Отрасль здравоохранения также применяет информационные технологии для усовершенствований, а именно в использовании информационных систем для ввода информации и дальнейшего формирования отчетности и документации. Также аналитика данных и машинное обучение позволяют извлекать необходимую и полезную информацию из медицинских исследований.

Обнаружение аномалий – важная задача, которая широко изучается во многих областях, связанных с интеллектуальным анализом данных. Применение в здравоохранении может выявить проблемы со вспышками заболеваемости, неправильным диагнозом, и, следовательно, неправильным лечением и т.д. С машинным обучением, мы можем обучить алгоритм классифицировать аномальные данные, изучая их характеристики.

Данные могут быть представлены в различных форматах: таблицы, изображения, аудио, видео, текст и т.д. Каждый тип представления данных имеет свой подход к обработке и анализу данных.

Сибирским государственным медицинским университетом были предоставлены деперсонализированные медицинские исследования. Которые были применены в данной работе.

Целью работы является повышение эффективности отрасли здравоохранения при анализе медицинских исследований путем разработки алгоритма обнаружения аномальных данных.

В рамках работы осуществляется обработка и подготовка медицинских исследований к анализу. Также предлагается алгоритм обнаружения аномальных данных в медицинских исследованиях.

Объект исследования - медицинские документы, сформированные на основании медицинских осмотров. Предметом исследования является процесс разработки алгоритмического и программного анализа результатов медицинских обследований.

Методы исследования — поиск литературы и источников, анализ информационных материалов, сравнение, консультация со специалистами.

В работе использованы различные методические (учебные) материалы и интернет ресурсы. Работа будет реализована на языке программирования «Python».



## ГЛАВА I. ОСНОВНАЯ ЧАСТЬ

### Обзор литературы

Многие сферы деятельности уже знакомы с понятием «большие данные». Медицинские, образовательные, финансовые и другие учреждения генерируют большие объемы информации. К сожалению, не все могут правильно работать с таким объемом данных, в большинстве случаев данные не сохраняются или хранятся без дальнейшего применения и анализа.

Международной консалтинговой компанией McKinsey определила 11 методов и приемов, применимых к большим данным, которые, в частности, включают в себя машинное обучение и визуализация аналитических данных [2]. Машинное обучение относится к построению алгоритмов, которые можно обучить с помощью математической статистики, численных методов, теории вероятностей, методов оптимизации, теории графов и других методов работы с цифровыми данными. Визуализация аналитических данных – это отображение исходной информации и результатов анализа данных в наиболее удобной для восприятия и интерпретации форме.

Алгоритмы обнаружения аномалий лежат в основе любой системы контроля качества данных. Под аномалией понимается любое отклонение от номинальных условий эксперимента.

В этой работе представлены основные теории машинного обучения, описаны алгоритмы, используемые для анализа медицинских данных, а также данные, используемые в этой работе.

Интеллектуальный анализ данных (Data Mining) - это область, которая позволяет выделить новую значимую информацию в большом количестве данных. Основная цель этой области -выявление явно невидимых взаимосвязей и неожиданных направлений изменений с использованием различных методов обработки данных.

Обнаружение аномалий -это поиск данных не соответствующих определенному понятию нормального, ожидаемого поведения. Процесс поиска аномалий состоит из определения области, которая представляет нормальное поведение, и определения аномалий в данных, которые не принадлежат этой нормальной области. Исходя из этого, аномалии определяются не по их характеристикам, а в сравнение с тем, что является нормой [1].

Одним из этапов этой работы является разделение медицинских исследований на два класса: нормальные данные и содержащие любые аномалии. Любая классификационная задача ставится следующим образом: существуют объекты, разделённые на классы. Имеется конечный набор, для которого известно, к каким классам принадлежат объекты этого набора. Этот набор называют обучающим (training set). Класс оставшихся объектов неизвестен. Нужно построить алгоритм, который назовёт имя класса, к которому относится свободный объект из исходного набора [7]. Для проверки работы обученного алгоритма необходимо подготовить тестовую выборку (validation set). Именно на этом наборе будет оцениваться точность классификации.

Весь процесс алгоритмов машинного обучения, как правило, делится на несколько логически следующих частей. На рисунке 1 приведена наиболее распространённая схема, содержащая основные этапы работы таких алгоритмов:

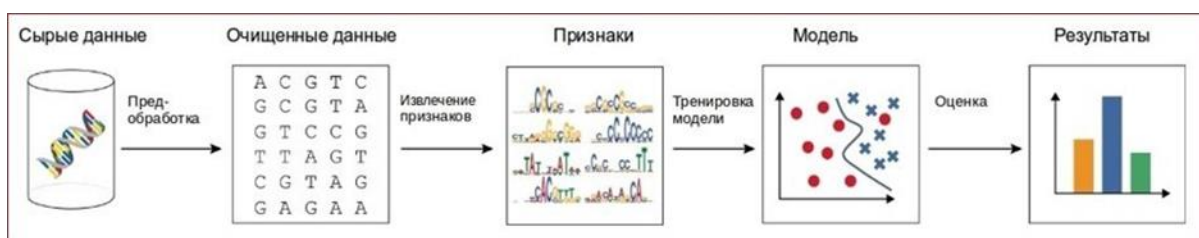


Рисунок 1 Классическая схема этапов машинного обучения

Существует множество наборов данных, содержащих медицинские данные, как здорового, так и патологического характера. После изучения нескольких примеров наборов данных, предпочтение было отдано наборам

данных, предоставленные Сибирским государственным медицинским университетом [3] и наборам данных загруженных с информационной платформы «Kaggle». Ключевым критерием выбора было количество данных, содержащееся в наборе, поскольку одной из особенностей алгоритмов обучения является потребность в максимально возможном количестве исходных данных.

Выбранные наборы данных в совокупности содержат немного больше 1 500 тысячи данных и находятся в свободном доступе [5].

## **1. Объект и методы исследования**

### **1.1 Методы машинного обучения**

Машинное обучение (Machine Learning) - это класс методов искусственного интеллекта, с не прямым решением проблемы, а обучением в процессе применения решений многих схожих проблем.

Машинное обучение используется, когда программирование высокопроизводительных алгоритмов затруднительно или невозможно. Качество и полнота обучающей выборки напрямую зависит от качества настройки системы.

Машинное обучение можно использовать для изучения закономерностей в данных, которые затем используются для выявления аномального поведения.

### **1.2 Обучение с учителем (Supervised learning)**

Обучение является подкатегорией машинного обучения и искусственного интеллекта. По мере того, как входные данные вводятся в модель, она корректирует свои характеристики до тех пор, пока модель не будет надлежащим образом подогнана, а ошибка будет минимизирована.

Обучение с учителем используется для задач классификации и прогнозирования. Обучение с учителем можно использовать для предсказания покупательского поведения, учитывая предыдущие модели. Контролируемое обучение помогает организациям решать разнообразные реальные проблемы, например, отделять в почтовом ящике спам в отдельную папку.

С учителем машина будет учиться быстрее, поэтому ее чаще используют в реальных задачах. Примеры задач: классификация и регрессия.

### **1.3 Обучение без учителя (Unsupervised learning)**

Обучение без учителя используется для исследовательского анализа данных, но не в качестве основного алгоритма. Алгоритм поначалу обучения получает не маркированные данные в качестве обучающего набора. В процессе работы алгоритм находит закономерности и схожие данные, и не связывает их.

Таким образом получая возможность найти интересные и неожиданные данные, узнать о данных больше, которые ранее не были обнаружены.

При обучении изучается набор данных, в которых обнаруживает скрытые взаимосвязи между различными переменными. Метод используется как математический процесс для систематического уменьшения излишек или структурирования данных по подобию.

Этот вариант имеет своё особое применение для компаний, которым необходимо, к примеру, объединить данные из разных источников, чтобы создать общую картину потребителей. К примеры таких задач можно отнести: кластеризация, уменьшение размерности и изучение правил ассоциации.

#### **1.4 Обучение с частичным привлечением учителя (Semi-Supervised learning)**

Это смешанное обучения с учителем и без учителя, в том случае, когда отсутствуют качественные данные. Алгоритм принимает данные как с меткой, так и без неё, и в большинстве без меток.

Доказано, что это обучение выдает точные результаты и может применяться к многим реальным задачам. Например, распознавание лиц в большом количестве различных людей на снимке.

Так же метод может использоваться для детектирования мошеннических действий на фоне привычной деятельности. Метод позволяет создавать алгоритмы, определяющие такие аномалии.

#### **1.5 Обучение с подкреплением (Reinforcement learning)**

При обучении с подкреплением вероятно связь с окружающий миром (например, выбрасывать бракованные продукты с ленты в корзину). Алгоритм подобен на обучение вашего питомца трюкам: вы дарите угощения в качестве награды, ежели ваш питомец осуществляет нужный вами трюк, в противном случае вы наказываете его. Вдобавок позволено допустить учиться самостоятельно. Этот способ особенно замысловатый и трудный, решает задачи

способом экспериментов и ошибок, реализовывается в мире и получении наград.

Одно из использований изучения с подкреплением — сортировка продуктов в магазинах. Роботы применяют обучение с подкреплением и глубокое обучение, дабы определить, как сильно придерживать объект, и какой захват представляется лучшим.

### **1.6 Глубокое обучение (Deep learning)**

Обучение использует программируемую нейронную сеть, которая позволяет машинам принимать точные решения без помощи человека, может реализовываться как без учителя, так и с подкреплением. Глубокое обучение структурирует алгоритмы слоями для создания искусственной нейронной сети, которая самостоятельно обучается и принимает решения. Подобное человеческим выводам.

Нейро-сети могут обрабатывать большое количество изображений за короткое время и извлекать функций, которые запоминает алгоритм.

К примерам глубокого обучения относится автоматическое распознавание речи.

### **1.7 Ансамблевые методы (Ensemble Methods)**

В ансамблевых методах используется идея комбинирования нескольких более слабых прогнозных алгоритмов для получения качественных прогнозов, чем каждый из алгоритмов может дать сама по себе. Это очень мощный класс, и поэтому он очень популярен. Например, алгоритм случайного леса - это способ ансамбля, что связывает множество деревьев решений, наученных с разными наборами данных. Итоги качества мониторингом случайного леса лучше, чем качество мониторингов одного дерева решений.

Методы ансамбля — это как способ уменьшить дисперсию и систематическую ошибку одной модели машинного обучения. Одна модель может быть точной при определенно заданных условиях, но неточной при

других условиях. С другой моделью относительная точность может быть обратной. Объединяя две модели, качество прогнозов балансируется.

Далее, в качестве методов, использованных для обнаружения аномалий, будут рассмотрены три алгоритма и их ансамбли.

## **2. Расчёты и аналитика**

### **2.1 Выбор программного обеспечения**

Среди языков программирования по работе с данными лидирующие позиции занимает «Python» [6]. «Python» обладает открытым исходным кодом, являются платформенно независимыми, необходимые пакеты легко устанавливаются, настраиваются и применяются.

В «Python» обладающий множеством готовых библиотек помимо анализа данных реализуется визуализация данных. «Python» является популярным, обладает большим количеством массивов документации, требуемой разработчикам для создания приложений. К недостаткам «Python» относят сложность отслеживания ошибок в коде.

Что касается простоты использования, то «Python» можно отнести к той группе языков, которые подходят на роль первого языка программирования. Преимуществом «Python» также являются особенности синтаксиса, стимулирующие разработчика писать удобно читаемый код.

#### **2.1.1 Используемые библиотеки «Python»**

В ходе реализации проекта были использованы некоторые «Python» библиотеки для работы с определенными типами данных и использования специальных функций, расширяющих возможности языка.

Библиотека «NumPy» – это библиотека, добавляющая поддержку многомерных массивов и матриц, а также математических операций над этими массивами.

Библиотека «pandas» – высокоуровневая «Python» библиотека, построенная поверх библиотеки «Python». Данная библиотека предоставляет возможности использования структур «Data Frame» и «Series». В ходе работы исходный набор данных хранился и обрабатывался в структуре «Data Frame», а также была использована функция «pandas.read\_csv» для считывания файла с исходными данными в структуру «Data Frame».



Библиотека «Matplotlib» представляет собой модуль-пакет для «Python», с помощью которого можно создавать рисунки и графики различных форматов. В ходе данной работы средства библиотеки были использованы для визуализации этапов работы и полученных результатов.

Библиотека «Scikit-learn» предоставляет возможности реализации ряда алгоритмов обучения, сокращения размерности данных, извлечения и отбора признаков. Несколько алгоритмов обучения, а также функции для извлечения и отбора признаков были использованы в текущей работе.

## **2.2 Загрузка и предварительный анализ данных**

Исходные данные хранятся в файлах с расширениями «.xls». Excel-файлы содержат в себе документы, содержащие информацию об медицинских исследованиях.

С помощью библиотеки «pandas» формируется объект «Data Frame», в который считываются транспонированные таблицы из файлов с расширением «.xls».

Элементы «Data Frame» подвергаются анализу: проверка полей на выбросы. Под выбросами понимают такие значения показателей, которые значительно отличаются от основного множества значений.

## **2.3 Разделение набора данных на тестовую и тренировочную выборки**

Для обеспечения проверки качества работы модели классификации возникает необходимость в разделении общего набора данных на тренировочную и тестовую выборки. Тренировочная выборка предназначена для обучения модели, тестовая выборка – для оценки качества работы модели.

Для разбиения набора данных на тренировочную и тестовую выборки используется метод «sklearn.model\_selection.train\_test\_split», в качестве параметра было указано, что тестовая выборка будет составлять 30% от общего набора данных.

### 3. Построение классификатора

#### 3.1 Выбор алгоритма классификации

В ходе работы для определения наилучшей алгоритмы классификации были рассмотрены следующие алгоритмы и их ансамбли:

- «Изолирующий лес» (Isolation Forest);
- «Одноклассовый метод опорных векторов» (One Class SVM);
- «Эллиптической огибающей» (Elliptic Envelope);

##### 3.1.1 Изолирующий лес (Isolation Forest)

Изолирующий лес (Isolation Forest) –это версия идеи случайного леса: простая и надёжная. Работает по принципу алгоритма дерева решений (рисунок 2). Он изолирует выбросы, выбирая функцию случайным образом из заданного набора функций, а затем случайным образом выбирая значение разделения между максимальным и минимальным значениями выбранной функции. Следовательно, когда лес случайных деревьев дает более короткие пути для конкретных точек, в таком случае они с большой вероятностью будут аномальными. Для каждого предмета нормальностью именуется среднее арифметическое значение глубины листьев, в которые он изолировался [8].

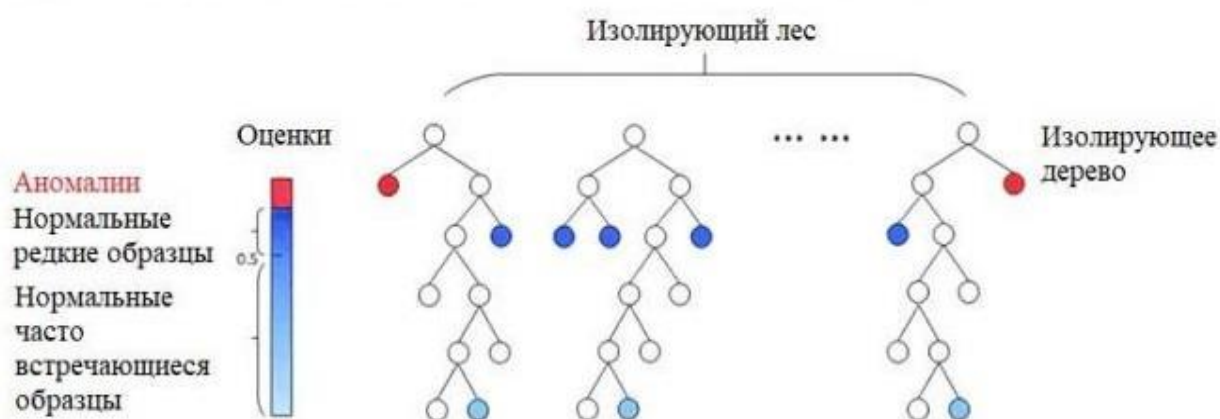


Рисунок 1 Схема выявления аномальности в алгоритме «изолирующий лес»

Логика метода такова, что при описанном методе построения деревьев выбросы будут упадать в листья, т.е. выбросы легче изолировать, а дерево будет строится пока произвольный объект не окажется в отдельном листе.

Ниже, на рисунке 3, показаны наблюдения нескольких наборов данных, полученные в результате анализа набора данных, идентифицированные этим методом как аномалии.

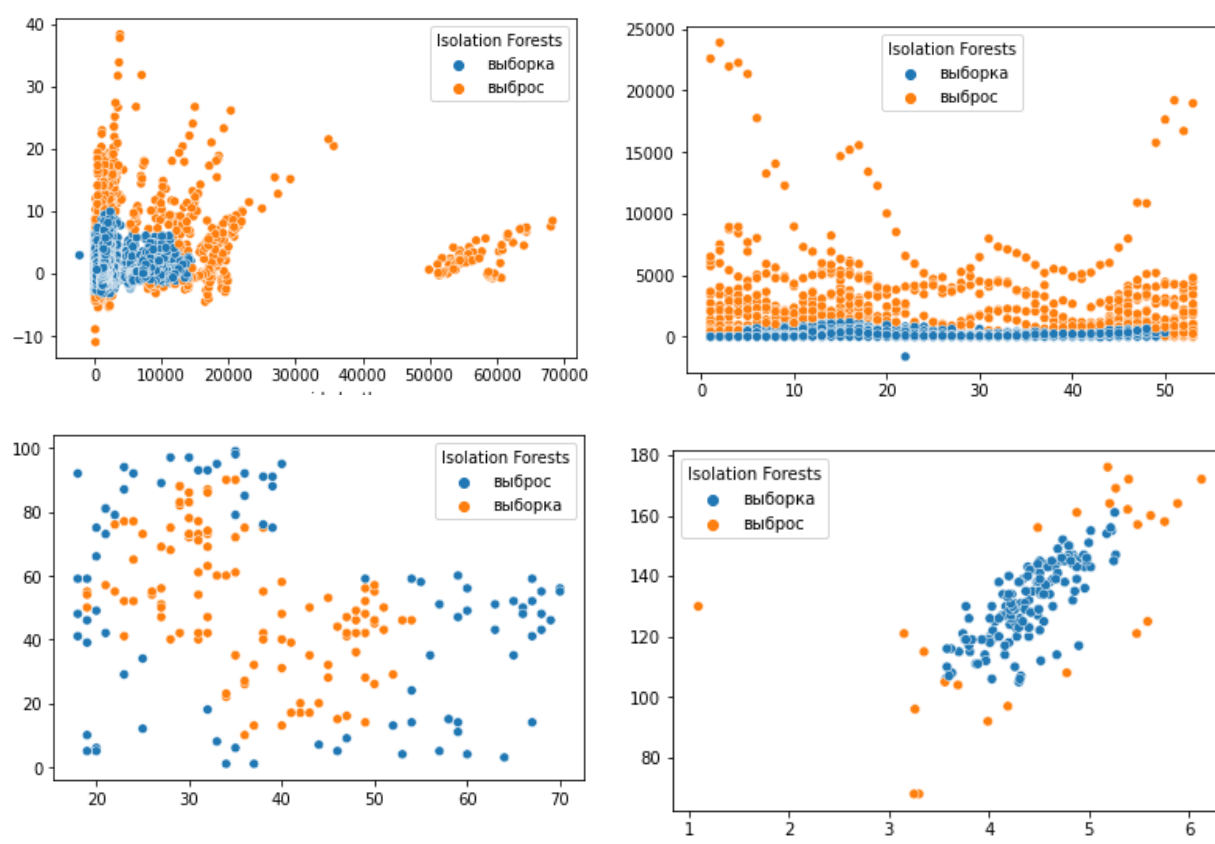


Рисунок 3 Isolation Forests

### 3.1.2 Одноклассовый метод опорных векторов (One Class SVM)

Одноклассовый метод опорных векторов (One Class SVM) — это один из классических алгоритмов, но для его обучения нам достаточно всего один класс [9].

В результате мы получаем границу, по одну сторону которой максимально плотно собраны наблюдения из выборки, а по другую будут находиться аномальные значения (рисунок 4).

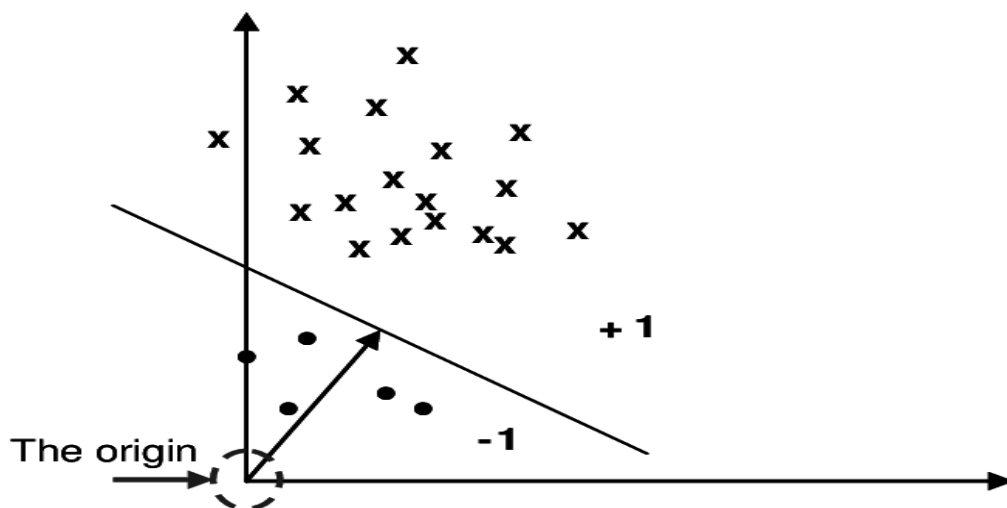


Рисунок 4 Классификация One Class SVM

На рисунке 5, приведены наблюдения нескольких наборов данных, полученные при анализе набора данных, идентифицированные данным методом как аномалии.

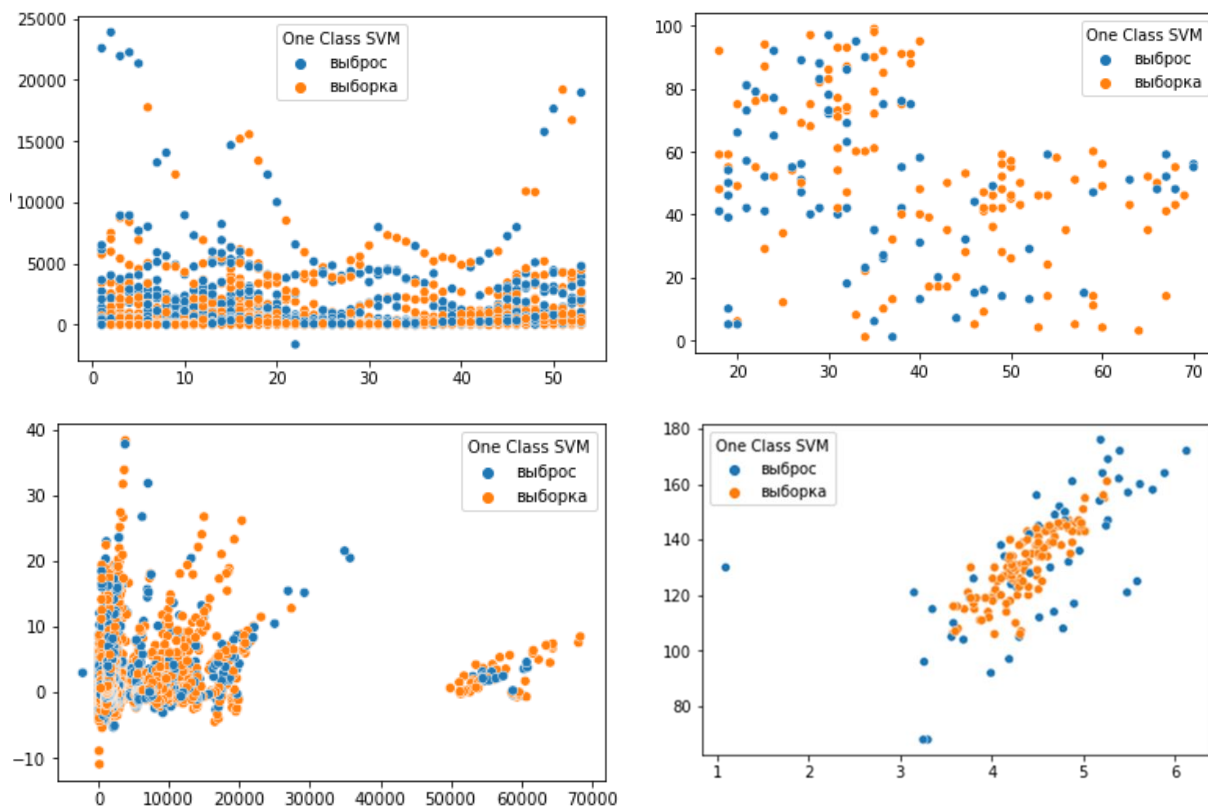


Рисунок 5 One Class SVM

### 3.1.3 Алгоритм эллиптической огибающей (Elliptic Envelope)

Алгоритм эллиптической огибающей (Elliptical Envelope) – создает воображаемую эллиптическую область вокруг набора данных. Данные,

попадающие в эту область, считаются нормальными данными, а все, что находится за пределами, возвращается как выбросы. Алгоритм работает лучше всего, если данные имеют гауссовское распределение [10].

На рисунке 6, приведены наблюдения нескольких наборов данных, полученные при анализе набора данных, идентифицированные данным методом как аномалии.

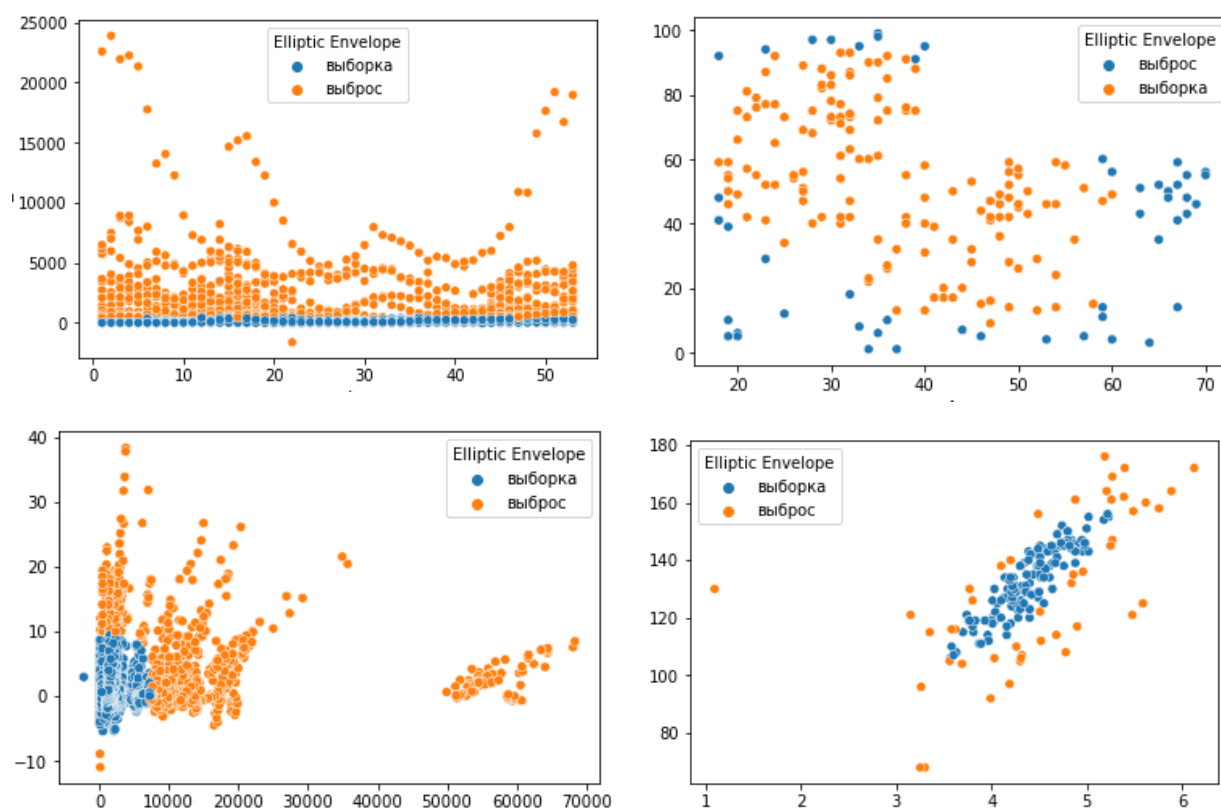


Рисунок 6 Выбросы Elliptic Envelope

### 3.1.4 Ансамбли алгоритмов

Ансамбли (Ensemble, Multiple Classifier System) это тип машинного обучения, в которой несколько моделей обучаются для решения одной и той же проблемы и объединяются для получения лучших результатов. Основная гипотеза состоит в том, что при правильном сочетании слабых моделей мы можем получить более точные и надежные модели [12]. Процесс построения ансамбля называется ансамблевым обучением (ensemble learning) [11]. Самый простой пример ансамбля в регрессии – усреднение нескольких алгоритмов (1):

$$a(x) = \frac{1}{n}(b_1(x) + \dots + b_n(x)) \quad (1)$$

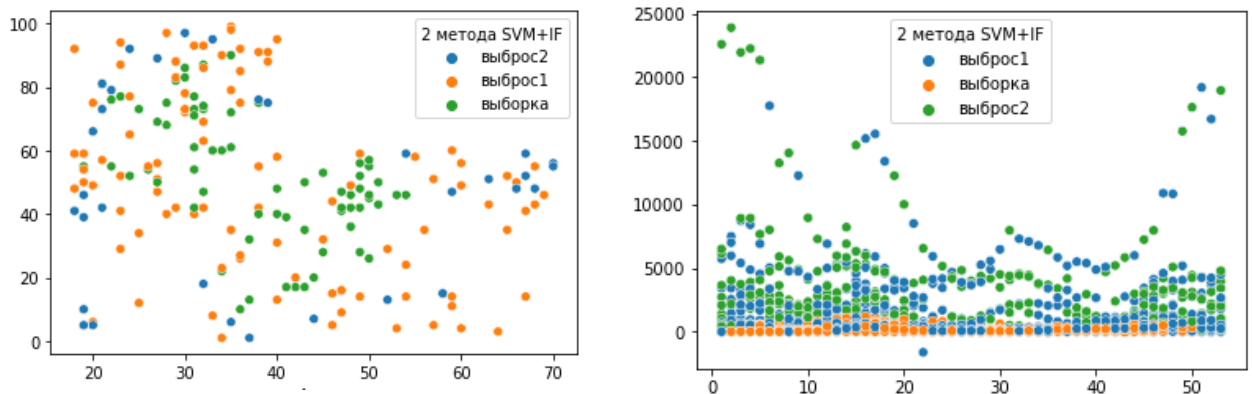
Алгоритмы, составляющие ансамбль ( $v(1) - bt$ ) называются базовыми алгоритмами (base learners). Если мы рассмотрим значения базовых алгоритмов на объекте как независимые случайные величины с одинаковым средним значением и одинаковой конечной дисперсией, то станет ясно, что случайная величина (2) имеет такое же среднее значение, но меньшую дисперсию [11]:

$$\begin{aligned} \xi &= \frac{1}{n}(\xi_1 + \dots + \xi_n) \\ \mathbf{E} \xi &= \frac{1}{n}(\mathbf{E} \xi_1 + \dots + \mathbf{E} \xi_n) = \mathbf{E} \xi_i \\ \mathbf{D} \xi &= \frac{1}{n^2}(\mathbf{D} \xi_1 + \dots + \mathbf{D} \xi_n) = \frac{\mathbf{D} \xi_i}{n} \end{aligned} \quad (2)$$

Для определения наилучшей алгоритмы классификации были рассмотрены следующие ансамбли алгоритмов:

### 1. Isolation Forest + One Class SVM

На рисунке 7, приведены наблюдения нескольких наборов данных, полученные при анализе набора данных, идентифицированные ансамблевым методом как аномалии.



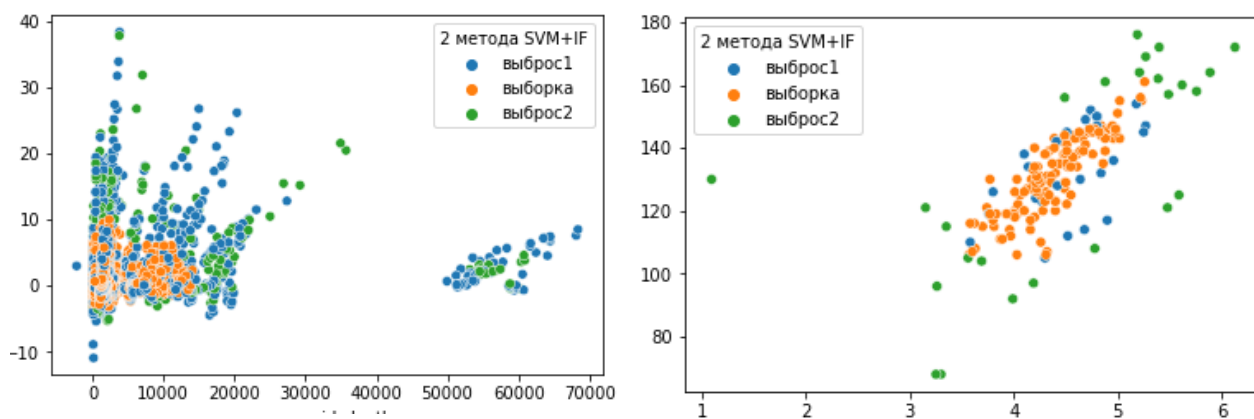


Рисунок 7 Результат работы ансамбля алгоритмов

## 2. One Class SVM + Elliptic Envelope

На рисунке 8, приведены наблюдения нескольких наборов данных, полученные при анализе набора данных, идентифицированные ансамблевым методом как аномалии.

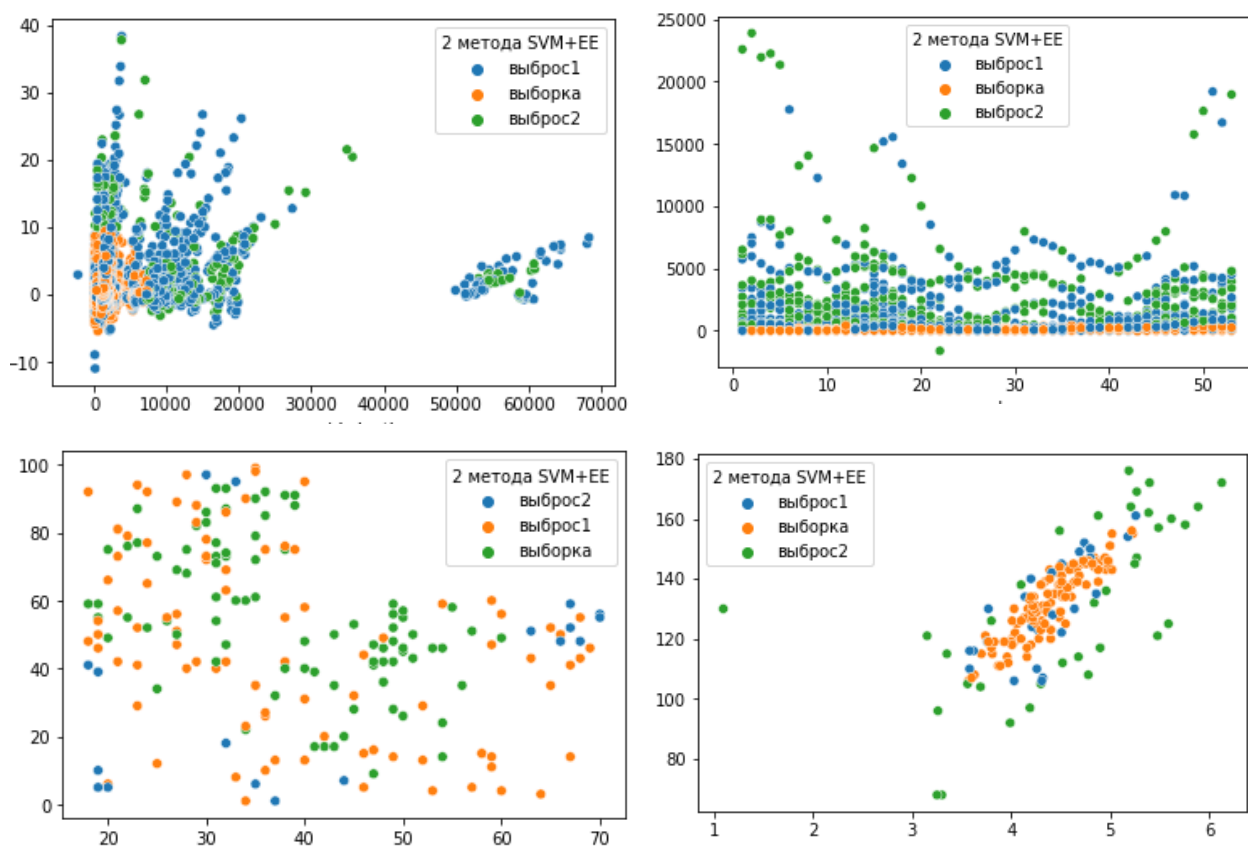


Рисунок 8 Результат работы ансамбля алгоритмов

### 3. Elliptic Envelope + Isolation Forest

На рисунке 9, приведены наблюдения нескольких наборов данных, полученные при анализе набора данных, идентифицированные ансамблевым методом как аномалии.

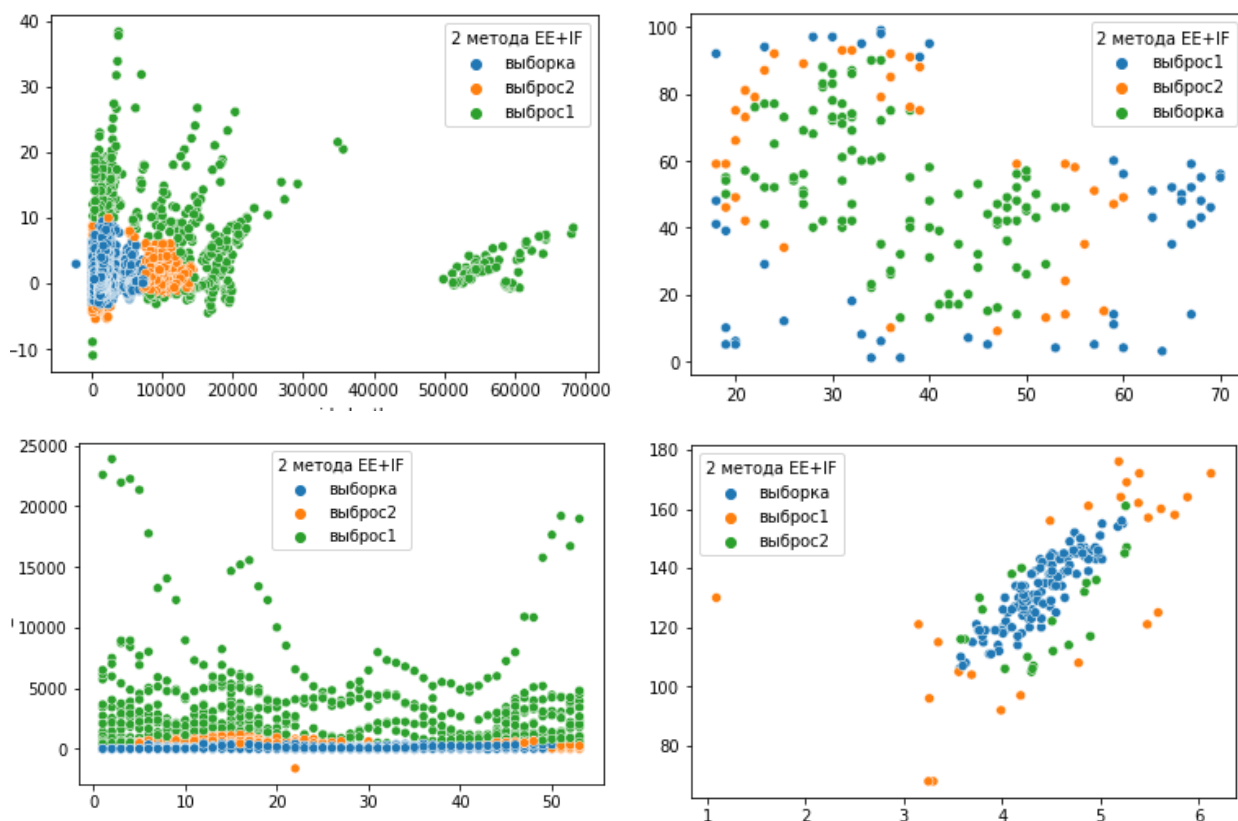


Рисунок 9 Результат работы ансамбля алгоритмов

### 4. Isolation Forest + One Class SVM + Elliptic Envelope

На рисунке 10, приведены наблюдения нескольких наборов данных, полученные при анализе набора данных, идентифицированные ансамблевым методом как аномалии.



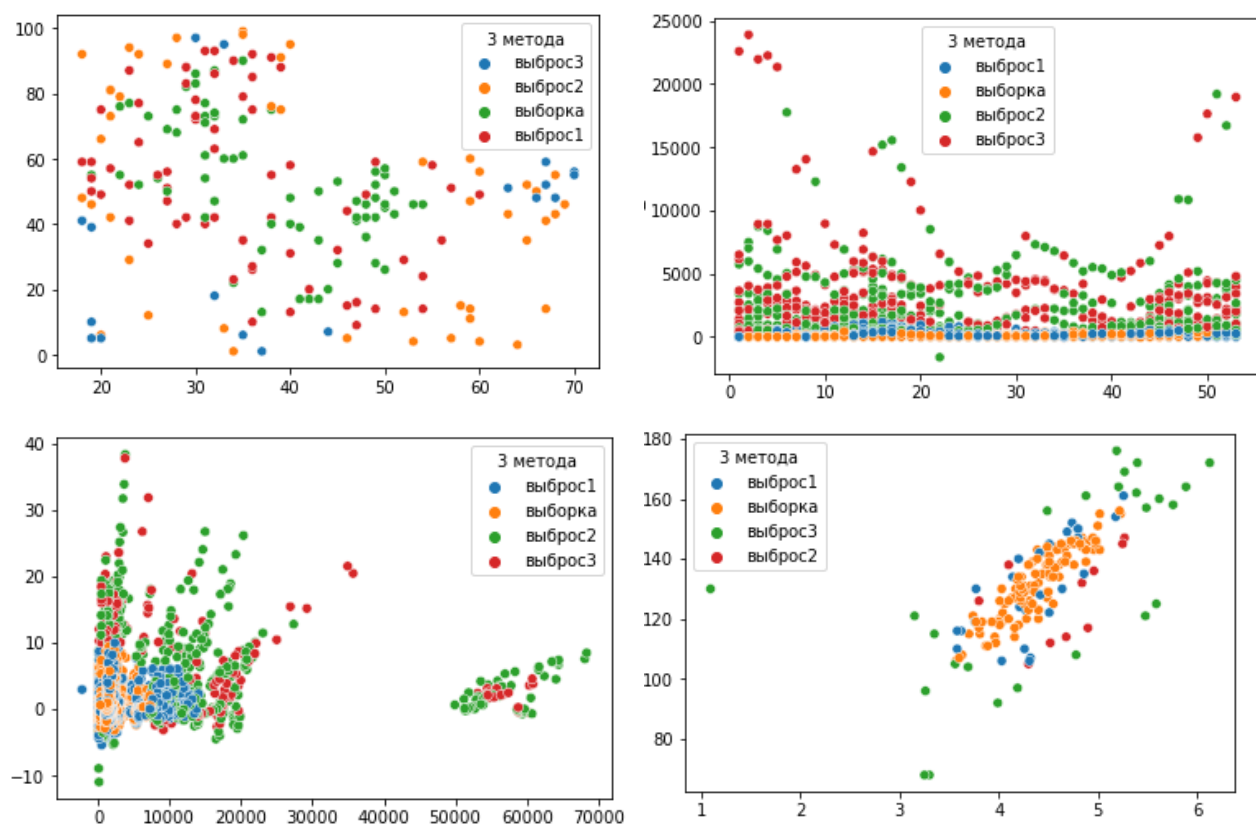


Рисунок 10 Результат работы ансамбля алгоритмов

#### 4. Результаты проведенного исследования

Со значениями параметров по умолчанию были построены семь моделей классификации. Результаты работы таких моделей, как «Isolation Forest», «Local Outlier Factor», «Elliptic Envelope» и их ансамбли, а именно, значения точности их предсказаний, представлены в таблице 1.

Таблица 1– Точность предсказания моделей классификации

№	EE + SVM	SVM + IF	SVM	IF+ EE	EE	IF+ EE+ SVM	IF
1	0,94	0,90	0,91	0,92	0,91	0,94	0,91
2	0,93	0,83	0,80	0,92	0,97	0,89	0,97
3	0,96	0,76	0,86	0,97	0,87	0,89	0,87
4	0,89	0,81	0,81	0,82	0,92	0,96	0,92
5	0,91	0,81	0,91	0,88	0,92	0,93	0,92
6	0,95	0,95	0,75	0,95	0,85	0,93	0,85
7	0,96	0,56	0,86	0,91	0,93	0,97	0,93
8	0,89	0,89	0,80	0,89	0,89	0,94	0,89
9	0,91	0,81	0,88	0,94	0,82	0,95	0,82
Среднее	<b>0,927</b>	<b>0,813</b>	<b>0,842</b>	<b>0,911</b>	<b>0,898</b>	<b>0,933</b>	<b>0,898</b>

Для обеспечения проверки качества работы алгоритмов используются фактические данные и матрица ошибок.

Матрица ошибок, представляет собой таблицу с четырьмя различными комбинациями, использующаяся для оценки эффективности алгоритмов. Количество правильных и неправильных прогнозов суммируется с фактическими значениями как истинные и ложные делятся по классам.

К примеру, дана выборка  $x_i(i=1, \dots, N, y_i)$  - метка класса  $i$ -го объекта,  $y_i \in \{1, 2, \dots, C\}$ , каждый объект которой относится к одному из  $C$  классов и классификатор  $a$ , который эти классы предсказывает.

Матрицей ошибок для такого классификатора называется следующая матрица (3):

$$M = \{m_{ij}\}_{i,j=0}^C, m_{ij} = \sum_{k=0}^N \mathbb{I}[a(x_k) = j] \mathbb{I}[y_k = i] \quad (3)$$

Такая матрица показывает, сколько объектов класса  $i$  были распознаны как объекты класса.

На рисунке 11 приведено изображение структуры матрицы ошибок 2x2. В случае бинарного разбиения, метка класса «у» получает значение +1 (положительный класс) или -1 (отрицательный). В качестве примера предположим, что было десять случаев, когда алгоритм классификации предсказывала «Да», в которых фактическое значение было «Да». Тогда число десять будет в верхнем левом углу квадранта истинного положительного результата. Это подводит нас к нескольким ключевым терминам:

		Actual Values	
		Yes	No
Predicted Values	Yes	True Positive	False Positive
	No	False Negative	True Negative

Рисунок 11 Матрица ошибок

1. Положительный (P): Наблюдение положительное (например, выброс).
2. Отрицательный (N): Наблюдение отрицательное (например, это не

выброс).

3. Истинно-положительный (TP): результат, при котором алгоритм правильно предсказывает положительный класс (4).

$$TP = \sum_{i=0}^n [a(x_i) = +1][y_i = +1] \quad (4)$$

4. Истинно-отрицательный (TN): результат, при котором алгоритм правильно предсказывает отрицательный класс (5).

$$TN = \sum_{i=0}^n [a(x_i) = -1][y_i = -1] \quad (5)$$

5. Ложно - положительный (FP): также называется ошибкой типа 1, результатом, когда алгоритм неверно предсказывает положительный класс, но на самом деле – отрицательный (6).

$$FP = \sum_{i=0}^n [a(x_i) = +1][y_i = -1] \quad (6)$$

6. Ложно - отрицательный (FN): также называется ошибкой 2-го типа, при котором алгоритм неверно предсказывает отрицательный класс, хотя на самом деле он положительный (7).

$$FN = \sum_{i=0}^n [a(x_i) = -1][y_i = +1] \quad (7)$$

Построим матрицу ошибок с помощью «sklearn.metrics.confusion\_matrix» и визуализируем ее. На рисунках 12,13,14,15,16 показаны результаты работы таких алгоритмов, как «Isolation Forest», «Local Outlier Factor», «Elliptic Envelope» и их ансамбли. Используя парные данные, заполнили матрицу ошибок.

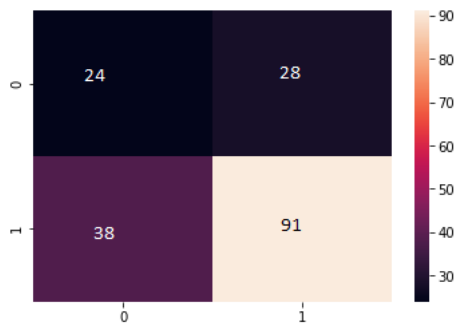


Рисунок 12 Elliptic Envelope + One Class SVM

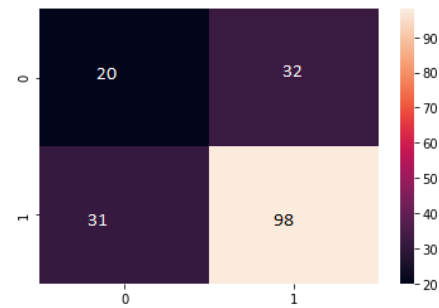


Рисунок 13 One Class SVM & One Class SVM + Isolation Forest

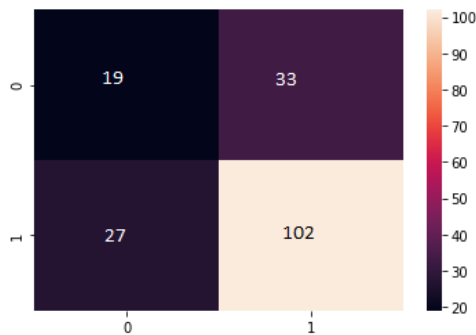


Рисунок 14 Isolation Forest+ Elliptic Envelope & Elliptic Envelope

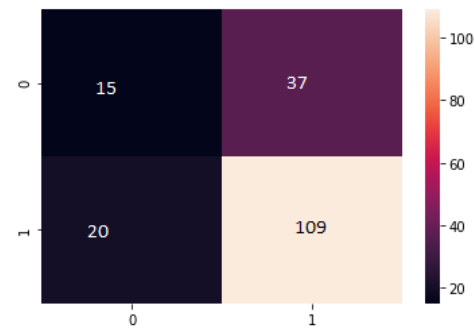


Рисунок 15 Isolation Forest+ Elliptic Envelope+ One Class SVM

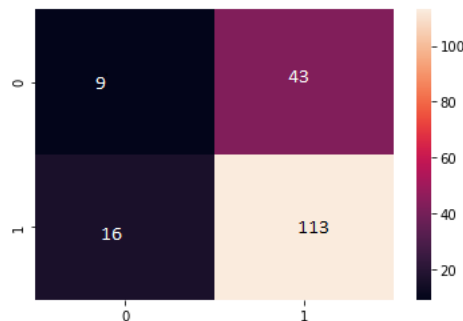


Рисунок 16 Isolation Forest

Исходя из полученных данных можно узнать о наших моделях несколько вещей: к примеру, по ансамблю алгоритмов «Elliptic Envelope + One Class SVM» в таблице видно, что у 62/181- выброс, тогда как на самом деле было 52/181 являются выбросами. Алгоритм имеет точность 115/181 или 63%. Нам нужно, чтобы алгоритм как можно точнее обнаружила выброс. Гораздо хуже, если алгоритм предскажет ложные данные, что приведет к ложному диагнозу.

Также для оценки качества алгоритмы используются значения таких

метрик качества как верность (accuracy), полнота (recall), точность (precision) и F1, которые представлены в таблице 2.

Таблица 2– Точность предсказания алгоритмов классификации для выбросов

Алгоритм	accuracy	recall	precision	f1
<b>IF+EE+SVM</b>	0.69	0.29	0.29	0.34
<b>EE + SVM</b>	0.64	0.46	0.39	0.42
<b>IF+ EE</b>	0.67	0.37	0.41	0.39

Исходя их анализа наибольшей средней точностью обладает ансамбль алгоритмов «Elliptic Envelope + One Class SVM». Значения всех метрик достаточно высоки, что говорит о высоком качестве работы алгоритма для поиска аномалий (выбросов).

Далее проведем оценку качества алгоритмов и их ансамблей на точность определения выборки по метрикам:

- верность (accuracy) формула (8)

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

- полнота (recall) формула (9)

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

- точность (precision) формула (10)

$$\text{Precision} = \frac{TP}{TP + FP} \quad (10)$$

- F1 формула (11)

$$f_1\text{-score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (11)$$

Результаты оценки качества приведены в таблице 3.

Таблица 3– Точность предсказания алгоритмов классификации для выборки

Алгоритм	accuracy	recall	precision	f1
<b>IF+EE+SVM</b>	0.69	0.84	0.75	0.79
<b>EE + SVM</b>	0.64	0.71	0.76	0.73
<b>IF+ EE</b>	0.67	0.79	0.76	0.77

Исходя из полученных данных, для решения этой задачи наибольшей средней точностью обладает ансамбль трёх алгоритмов «Isolation Forest + Elliptic Envelope + One Class SVM», так же можно выделить ансамбль «Isolation Forest + Elliptic Envelope». Значения всех метрик достаточно высоки, что говорит о высоком качестве работы алгоритма.

## **Заключение по разделу**

Выбор алгоритма поиска аномалий зависит в первую очередь от первоначальной задачи, данных и доступной исходной информации. Как было продемонстрировано в этой работе, перед исследованием возникает потребность предварительной обработки данных, по крайней мере, в области, в которой ожидаются аномальные данные. Обученная сеть не всегда способна правильно отображать поведение системы. Это связано с тем, что для обучения на примерах использования требуется большой объем данных, в то время как нейронная сеть не может предсказать ситуации, которые отсутствовали в исторических данных.

Основываясь на сравнительный анализ, наиболее стабильным и общим алгоритмом как для определения аномалий, так и для определения выборки остаётся ансамбль алгоритмов «Isolation Forest + Elliptic Envelope». Поскольку акцент работы направлен на определение аномалий, стоит выделить ансамбль алгоритмов «Elliptic Envelope + One Class SVM + Isolation Forest» для решения этой задачи, и именно он считается лучшим.



## **ГЛАВА II. ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ**

### **Введение**

Экономическая часть ВКР предназначена для комплексного описания и анализа финансово - экономических аспектов выполненной работы. Целью данного раздела является планирование и формирование бюджета научных исследований, определение ресурсной, финансовой, бюджетной, социальной и экономической эффективности исследования.

Целью настоящей работы является разработка алгоритма для выявления аномалий в медицинских наборах данных путем применения ряда алгоритмов и их ансамблей. Разрабатываемое решение предназначено для использования в медицинских учреждениях для улучшения качества анализа и постановки верного диагноза пациенту.

В качестве потенциальных потребителей результатов научно-исследовательского проекта могут выступать организации здравоохранения, медицинский персонал, а также представители научного сообщества, заинтересованные в продолжении исследования или в использовании его технических результатов.

### **1. Организация и планирование работ**

При организации процесса был определен полный перечень необходимых работ, а также их исполнители и рациональная продолжительность. В качестве структуры, показывающей необходимые данные, был использован линейный график работ, представленный в таблице 4.

Таблица 4 – Перечень работ и продолжительность их выполнения

Этапы работы	Исполнители	Загрузка исполнителей
Постановка целей и задач	НР	НР – 100%
Составление и утверждение ТЗ	НР, И	НР – 100% И – 10%
Разработка календарного плана	НР, И	НР – 100% И – 10%
Анализ исследуемой области	И	И – 100%
Подготовка данных для анализа	И	И – 100%
Разработка алгоритма определения аномалий	И	И – 100%
Тестирование алгоритмов	И	И – 100%
Оценка эффективности полученных результатов	НР, И	НР – 50% И – 100%
Выполнение разделов «Финансовый менеджмент» и «Социальная ответственность»	И	И – 100%
Оформление пояснительной записки	И	И – 100%
<i>Примечание: НР – научный руководитель, И – инженер</i>		

## 2. Продолжительность этапов работ

Для расчета продолжительности этапов работ был выбран экспертный опытно-статистический метод. Определение вероятных (ожидаемых) значений продолжительности работ было выполнено по формуле (12):

$$t_{ож} = \frac{3 \cdot t_{min} + 2 \cdot t_{max}}{5}, \quad (12)$$

где:

$t_{min}$  – минимальная продолжительность работы, дн.;

$t_{max}$  – максимальная продолжительность работы, дн.

Для построения линейного графика необходимо рассчитать длительность этапов в рабочих днях, а затем перевести ее в календарные дни. Расчет продолжительности этапа в рабочих днях был рассчитан по формуле (13):

$$T_{РД} = \frac{t_{ож}}{K_{ВН}} \cdot K_{Д}, \quad (13)$$

где:

$K_{ВН}$  – коэффициент выполнения работ, учитывающий влияние внешних

факторов на соблюдение предварительно определенных;

$K_{Д}$  – коэффициент, учитывающий дополнительное время на компенсацию непредвиденных задержек и согласование работ. Примем  $K_{Д} = 1,1$ .

Формула расчета продолжительности этапа в календарных днях (14):

$$T_{КД} = T_{РД} \cdot T_{К}, \quad (14)$$

где:

$T_{К}$  – коэффициент календарности, позволяющий перейти от длительности работ в рабочих днях к их аналогам в календарных днях, и рассчитываемый по формуле (15):

$$K = \frac{T_{КАЛ}}{T_{КАЛ} - T_{ВД} - T_{ПД}}, \quad (15)$$

где:

$T_{КАЛ}$  – календарные дни, дн.;

$T_{ВД}$  – выходные дни, дн.;

$T_{ПД}$  – праздничные дни, дн.

При шестидневной рабочей неделе в 2021 году коэффициент календарности равен:

$$T_{К} = \frac{365}{365 - 14 - 42} = 1,22$$

Полученные результаты трудозатрат на выполнение проекта отображены в таблице 5, а линейный график работ – на рисунке 17.

Таблица 5 – Трудозатраты на выполнение проекта

Этап	Исполнители	Продолжительность работ, дни			Трудоемкость работ по исполнителям, чел-дн.			
		$t_{min}$	$t_{max}$	$t_{ож}$	$T_{РД}$		$T_{КД}$	
					НР	И	НР	И
Постановка целей и задач	НР	1	2	1,4	1,54	–	1,88	–
Составление и утверждение ТЗ	НР, И	4	6	4,8	5,28	0,53	6,44	0,64
Разработка календарного плана	НР, И	1	2	1,4	1,54	0,15	1,88	0,19
Анализ исследуемой области	И	6	8	6,8	–	7,48	–	9,13
Подготовка данных для анализа	НР, И	12	18	14,4	11,09	15,84	13,53	19,32
Разработка алгоритма определения аномалий	И	24	30	26,4	–	29,04	–	35,43
Тестирование алгоритмов	И	10	15	12	–	10,82	–	13,2
Оценка эффективности полученных результатов	НР, И	7	11	8,6	6,31	7,08	7,7	8,64
Выполнение разделов «Финансовый менеджмент» и «Социальная ответственность»	И	2	4	2,8	–	3,08	–	3,76
Оформление пояснительной записки	И	12	18	14,4	–	15,84	–	19,32
<b>Итого:</b>				<b>93</b>	<b>25,76</b>	<b>89,86</b>	<b>31,43</b>	<b>109,63</b>

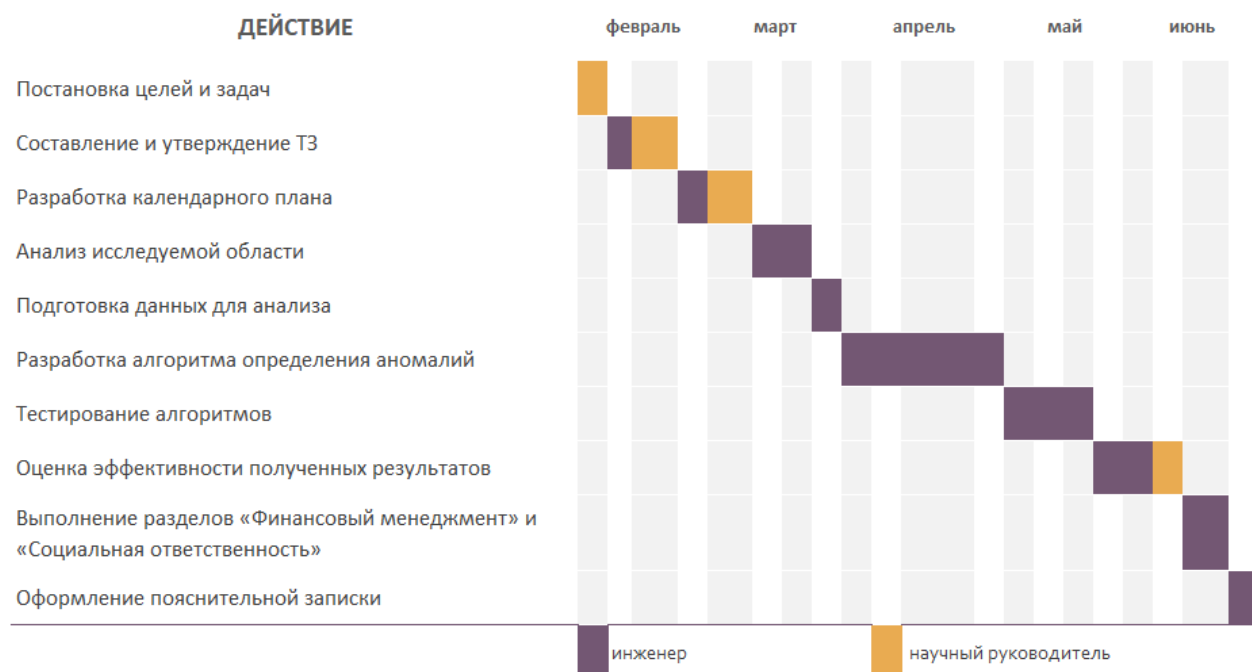


Рисунок 17 – Линейный график работ:

оранжевым цветом выделены работы научного руководителя, сиреневым –  
инженера

### 3. Расчет сметы затрат на выполнение проекта

В состав затрат на создание проекта включается величина всех расходов, необходимых для реализации комплекса работ, составляющих содержание данной разработки. Расчет сметной стоимости ее выполнения производился по следующим статьям затрат:

- материалы и покупные изделия;
- заработная плата;
- социальный налог;
- расходы на электроэнергию (без освещения);
- амортизационные отчисления;
- прочие (накладные расходы) расходы.

### 4. Расчет заработной платы

Данная статья расходов включает заработную плату научного руководителя и исполнитель проекта, а также премии, входящие в фонд заработной платы.

Среднедневная тарифная заработная плата была рассчитана по

формуле (16):

$$ЗП_{дн-г} = \frac{МО}{25} \quad (16)$$

где:

*МО* – месячный оклад сотрудника.

Учитывая, что в 2021 году 300 рабочих дней при шестидневной рабочей неделе, в каждом месяце имеется около 25 рабочих дней.

Для перехода от тарифной (базовой) суммы заработка сотрудника, связанной с участием в проекте, к соответствующему полному заработку необходимо умножить на интегральный коэффициент  $K_{и}$ , который равен 1,699 при шестидневной рабочей неделе. Расчет затрат на полную заработную плату приведен в таблице 6.

Таблица 6 – Затраты на заработную плату

Исполнитель	Оклад, руб./мес.	Среднедневная ставка, руб./раб. дн.	Затраты времени, раб. дни	Фонд заработной платы, руб.
НР	33 664	1 342,1	31	70 687,3
И	15 470	616,75	90	94 307,6
<b>Итого:</b>				<b>1644,9</b>

## 5. Расчет затрат на социальный налог

Затраты на единый социальный налог (ЕСН), включающий в себя отчисления в пенсионный фонд, на социальное и медицинское страхование, составляют 30 % от полной заработной платы по проекту. Соответственно, для текущего проекта затраты на социальный налог равны:

$$C_{соц.} = 164\,994,9 \cdot 0,3 = 49\,498,5 \text{ руб.}$$

## 6. Расчет затрат на электроэнергию

Данный вид расходов включает в себя затраты на электроэнергию, потраченную в ходе выполнения проекта на работу используемого оборудования. Расчет затрат на электроэнергию рассчитывается по формуле (17):

$$C_{\text{эл.об.}} = P_{\text{об}} \cdot t_{\text{об}} \cdot C_{\text{э}}, \quad (17)$$

где:

$P_{\text{об}}$  – мощность, потребляемая оборудованием, кВт;

$C_{\text{э}}$  – тариф на 1 кВт·час;

$t_{\text{об}}$  – время работы оборудования, час. Для ТПУ  $C_{\text{э}} = 6,59$  руб./кВт·час (с НДС).

Время работы оборудования вычисляется на основе итоговых данных таблицы 2 для инженера ( $T_{\text{рд}}$ ) из расчета, что продолжительность рабочего дня равна 8 часов формула (18):

$$t_{\text{об}} = T_{\text{рд}} \cdot K_t, \quad (18)$$

где  $K_t$  – коэффициент использования оборудования по времени, равный отношению времени его работы в процессе выполнения проекта к  $T_{\text{рд}}$ .

Мощность, потребляемая оборудованием, определяется по формуле (19):

$$P_{\text{об}} = P_{\text{ном.}} \cdot K_c, \quad (19)$$

где:

$P_{\text{ном.}}$  – номинальная мощность оборудования, кВт;

$K_c$  – коэффициент загрузки, зависящий от средней степени использования номинальной мощности.

Расчет затрат на электроэнергию для технологических целей приведен в таблице 7.

Таблица 7 – Расчет затрат на электроэнергию

Наименование оборудования	Время работы оборудования, час	Потребляемая мощность, кВт	Затраты, руб.
Ноутбук	701 · 0,8	0,045	167,4



### 6.1. Расчет амортизационных расходов

В статье «Амортизационные отчисления» рассчитывается амортизация используемого оборудования за время выполнения проекта. Для этого была использована формула (20):

$$C_{AM} = \frac{H_A \cdot C_{ОБ} \cdot t_{pф} \cdot n}{F_d}, \quad (20)$$

где:

$H_A$  – годовая норма амортизации единицы оборудования;

$C_{ОБ}$  – балансовая стоимость единицы оборудования с учетом ТЗР;

$F_d$  – действительный годовой фонд времени работы соответствующего оборудования;

$t_{pф}$  – фактическое время работы оборудования в ходе выполнения проекта;

$n$  – число задействованных однотипных единиц оборудования.

При реализации проекта был использован ПК, и учитывая, что при шестидневной рабочей неделе в 2021 году имеется 300 рабочих дней, примем  $F_d = 300 \cdot 8 = 2400$  часа. Также примем  $H_A$  равным 0,4, так как срок амортизации ПК составляет порядка 2,5 лет.

Таким образом, при реализации проекта амортизация оборудования (ноутбука) составила:

$$C_{AM} = \frac{0,4 \cdot 89990 \cdot 701 \cdot 1}{2400} = 10\,513,8 \text{ руб.}$$

### 6.2. Расчет прочих расходов

В статье «Прочие расходы» отражены расходы на выполнение проекта, которые не учтены в предыдущих статьях. Их следует принять равными 10% от суммы всех предыдущих расходов (21):

$$C_{проч.} = (C_{мат} + C_{зн} + C_{соц} + C_{эл.об.} + C_{ам}) \cdot 0,1, \quad (21)$$

Таким образом, прочие расходы на реализацию настоящего проекта составили:

$$C_{\text{проч.}} = (2\,500 + 164\,994,9 + 49\,498,5 + 167,4 + 10\,513,8) \cdot 0,1 = 22\,767,5 \text{ руб.}$$

## 7. Расчет общей себестоимости разработки

Проведя расчет по всем статьям сметы затрат на разработку, можно определить общую себестоимость проекта «Алгоритмическое и программное обеспечение поиска аномалий в результатах медицинских исследований» (таблица 8).

Таблица 8 – Расчет себестоимости разработки проекта

Статья затрат	Условное обозначение	Сумма, руб.
Материалы и покупные изделия	$C_{\text{мат}}$	2 500
Основная заработная плата	$C_{\text{зп}}$	164 994,9
Отчисления в социальные фонды	$C_{\text{соц}}$	49 498,5
Расходы на электроэнергию	$C_{\text{эл.}}$	167,4
Амортизационные отчисления	$C_{\text{ам}}$	10 513,8
Прочие расходы	$C_{\text{проч}}$	22 767,5
<b>Итого:</b>		<b>2502</b>

### 7.1. Расчет прибыли

Так как получить данные для применения методов для расчета прибыли не представляется возможным, поэтому прибыль была принята в размере 20 % от полной себестоимости проекта и составила.

$$C_{\text{пр}} = 250\,442 \cdot 0,2 = 50\,088,4 \text{ руб.}$$

### 7.2. Расчет НДС

НДС составляет 20% от суммы затрат на разработку и прибыли:

$$C_{\text{НДС}} = (250\,442 + 50\,088,4) \cdot 0,2 = 60\,106,1 \text{ руб.}$$

### **7.3. Цена разработки НИР**

Цена равна сумме полной себестоимости, прибыли и НДС:

$$C_{НИР(КР)} = 250\,442 + 50\,088,4 + 60\,106,1 = 360\,636,5 \text{ руб.}$$

### **8. Оценка экономической эффективности проекта**

Предметом реализованного проекта является программное обеспечение, предназначенное для поиска аномалий в результатах медицинских исследований. На данный момент существует ряд алгоритмов для анализа медицинских данных. Однако, ни один из исследованных алгоритмов в рамках настоящего проекта не предоставляет удовлетворительный результат анализа.

Потенциальная косвенная экономическая эффективность реализованного проекта заключается в снижении нагрузки на медицинские учреждения. Ожидается, что врач, получив результаты анализа аномалий, вероятно будет иметь точное представление о постановке диагноза. Однако разработанный алгоритм не был внедрен, и оценить изменение затрачиваемого времени врачом в медицинском учреждении не представляется возможным.

В рамках выполненной работы не была осуществлена экспертная оценка эксплуатационных издержек на реализованный проект, поэтому оценка данного экономического результата невозможна.

#### **Заключение по разделу**

Подсчитанные затраты и преимущества от выполнения проекта позволяют доказать экономическую целесообразность проекта при благополучном исходе. В рамках работы были подсчитаны приблизительные затраты на выполнение проекта, которые в сумме составили 360 636,5 рублей. Основная часть затрат приходится на заработную плату исполнителей проекта.

## **ГЛАВА III. СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ**

### **Введение**

Данная работа посвящена разработке алгоритма выявления аномалий по медицинским данным. Данный алгоритм может найти свое применение в системах поддержки принятия решений врача, которые смогут помочь специалисту сократить время, а также увеличить точность принятия решений о способе лечения пациентов на основе схожих случаев.

Работа выполнялась с использованием ЭВМ на базе Отделения информационных технологий Инженерной школы информационных технологий и робототехники Томского политехнического университета (ТПУ), данные для работы были предоставлены Отделением инфекционных заболеваний Сибирского государственного медицинского университета.

В разделе будут рассмотрены опасные и вредные факторы, оказывающие влияние на производственную деятельность инженера-программиста. Исследовано рабочее место программиста и помещение, в котором оно находится. Разработка осуществлялась в компьютерном классе Кибернетического центра ТПУ. Рассмотрены воздействия объекта исследования на окружающую среду, правовые и организационные вопросы, а также мероприятия в чрезвычайных ситуациях.

### **1. Правовые и организационные вопросы обеспечения безопасности**

#### **1.1. Специальные правовые нормы трудового законодательства**

Основным документом, регулирующим отношения между работником и работодателем, является Трудовой Кодекс Российской Федерации (ТК РФ) [13].

Согласно трудовому кодексу РФ, нормальная продолжительность рабочего времени не может превышать 40 часов в неделю. Также работнику в течение рабочего дня (смены) должен предоставляться, перерыв на отдых и питание продолжительностью от 30 минут до 2 часов, который не включается в рабочее время. Если продолжительность ежедневной работы не превышает

4 часов, то указанный перерыв может не предоставляться. Всем работникам предоставляются выходные дни (еженедельный непрерывный отдых).

Федеральный закон «О специальной оценке условий труда» [14] регламентирует проведение специальной оценки в случае, если деятельность работников предприятия предусматривает непрерывную работу за компьютерными системами. По проведенной оценке, устанавливаются гарантии и компенсации работникам согласно ТК РФ.

## 1.2. Эргономические требования к рабочему месту оператора ПЭВМ

Главными элементами рабочего места программиста являются стол, кресло, дисплей, клавиатура и мышь. Основным рабочим положением является положение сидя. Нормативными положениями ГОСТ 12.2.032-78 ССБТ, ГОСТ 12.2.033-78 ССБТ и рядом других предъявляются определенные требования к оснащению рабочего места, предусматривающего длительную работу за персональным компьютером (Таблица 9).

Таблица 9 – Нормы оборудования рабочих мест с ПЭВМ

Параметр		Значение
Высота перегородок, разделяющих рабочие места		Не менее 1,5 метров
Стол	Ширина рабочей поверхности	От 80 до 140 см
	Глубина рабочей поверхности	От 80 до 100 см
	Высота рабочей поверхности	72,5 см
Расстояние от глаз до монитора		От 60 до 70 см
Расстояние клавиатуры от края стола		От 10 до 30 см
Стул	Ширина поверхности	От 40 см
	Глубина поверхности	От 40 см
	Регулировка высоты поверхности	От 40 до 50 см
	Угол наклона вперед	До 15 градусов
	Угол наклона назад	До 5 градусов
	Высота опорной поверхности спинки	30 плюс/минус 2 см
	Ширина опорной поверхности спинки	От 38 см
	Радиус кривизны горизонтальной плоскости спинки	40 см
	Угол наклона спинки в вертикальной плоскости	Плюс/минус 30 градусов
	Регулировка расстояния спинки от переднего края сидения	От 26 до 40 см
	Длина подлокотников	От 25 см
	Ширина подлокотников	От 5 до 7 см
	Регулировка подлокотников по высоте над сиденьем	23 плюс/минус 3 см

	Внутреннее расстояние между подлокотниками	От 35 до 50 см
Подставка для ног	Ширина	От 30 см
	Глубина	От 40 см
	Регулировка по высоте	До 15 см
	Угол наклона опорной поверхности	До 20 градусов
	Высота бортика по переднему краю	1 см

Работающий на ПЭВМ должен сидеть прямо, опираясь в области нижнего края лопаток на спинку кресла, не сутулясь, с небольшим наклоном головы вперед (до 5 – 7 градусов). Предплечья должны опираться на поверхность стола, снимая статическое напряжение плечевого пояса и рук.

Рабочие места в компьютерном классе Кибернетического центра (КЦ) ТПУ отвечают данным требованиям.

## 2. Производственная безопасность

### 2.1. Анализ вредных и опасных факторов, которые может создать объект исследования

Опасные и вредные производственные факторы, выявленные в рамках настоящей работы, представлены в таблице 10.

Таблица 10 – Опасные и вредные производственные факторы

Факторы (ГОСТ 12.0.003-2015)	Этапы работ		Нормативные документы
	Разработка	Эксплуатация	
Отклонение показателей микроклимата	+	+	СанПиН 2.2.4.548-96 «Гигиенические требования к микроклимату производственных помещений»
Превышение уровня шума	+	+	СанПиН 2.2.4.3359-16 «Санитарно-эпидемиологические требования к физическим факторам на рабочих местах».
Отсутствие или недостаток естественного света	+	+	СП 52.13330.2016 «Естественное и искусственное освещение». Актуализированная редакция СНиП 23-05-95.

Повышенное значение напряжения в электрической цепи, замыкание которой может произойти через тело человека	+	+	Защитное заземление, зануление, ГОСТ 12.1.030–81 ССБТ. ГОСТ Р 12.1.019-2009 ССБТ. Электробезопасность. Общие требования и номенклатура видов защиты.
--	---	---	---

## **2.2. Обоснование мероприятий по снижению воздействия вредных и опасных факторов**

### **2.2.1. Отклонение показателей микроклимата**

Для обеспечения установленных норм микроклиматических параметров и чистоты воздуха на рабочих местах и в помещениях применяют вентиляцию. Обще обменная вентиляция используется для обеспечения в помещениях соответствующего микроклимата. Периодически должен вестись контроль за влажностью воздуха.

В летнее время при высокой уличной температуре должны использоваться системы кондиционирования. В холодное время года предусматривается система отопления. Для отопления помещений используются водяные системы центрального отопления.

Для категории тяжести работ 1а (работы с интенсивностью энергозатрат до 120 ккал/ч (до 139 Вт), производимые сидя и сопровождающиеся незначительным физическим напряжением) оптимальные и допустимые показатели температуры, относительной влажности и скорости движения воздуха в рабочей зоне производственных помещений должны соответствовать значениям, представленным в таблице 11 «Гигиенические требования к микроклимату производственных помещений» [17].

Таблица 11 – Оптимальные нормы микроклимата

Период года	Температура, °С	Температура поверхностей, °С	Относительная влажность, %	Скорость движения воздуха, м/с
Холодный	22-24	21-25	40-60	0,1
Теплый	23-25	22-26	40-60	0,1

Допустимые микроклиматические условия установлены по критериям

допустимого теплового и функционального состояния человека на период 8-часовой рабочей смены. Они устанавливаются в случаях, когда по технологическим требованиям, технически и экономически обоснованным причинам не могут быть обеспечены оптимальные величины. В таблице 12 представлены допустимые величины показателей микроклимата на рабочих местах.

Таблица 12 – Допустимые величины показателей микроклимата

Период года	Температура воздуха, °С		Относительная влажность воздуха, %	Скорость движения воздуха, не более, м/с	
	Диапазон ниже оптимальных величин	Диапазон выше оптимальных величин		при температуре воздуха ниже оптимальной	при температуре воздуха выше оптимальной
Холодный	20,0 – 21,9	24,1 – 25,0	15 – 75	0,1	0,1
Теплый	21,0 – 22,9	25,1 – 28,0	15 - 75	0,1	0,2

При обеспечении допустимых величин микроклимата на рабочих местах перепад температуры воздуха по высоте должен быть не более 3°С, перепад температуры воздуха по горизонтали, а также ее изменения в течение смены не должны превышать – 4°С. При этом абсолютные значения температуры воздуха не должны выходить за пределы оптимальных величин.

При температуре воздуха на рабочих местах 25°С и выше максимально допустимые значения относительной влажности воздуха не должны выходить за следующие пределы:

- 70% – при температуре воздуха 25°С,
- 65% – при температуре воздуха 26°С,
- 60% – при температуре воздуха 27°С,
- 55% – при температуре воздуха 28°С



Для обеспечения установленных норм микроклиматических параметров и чистоты воздуха на рабочих местах и в помещениях применяется вентиляция, периодически ведется контроль влажности воздуха. В летнее время при высокой уличной температуре используется система кондиционирования. В холодное время года предусматривается система отопления. Для отопления помещений используются водяные системы центрального отопления.

В компьютерной аудитории № 418 КЦ ТПУ поддерживается температура, равная 19–20 С°, при относительной влажности 55–58%, что удовлетворяет требованиям СанПиН 2.2.4.548-96.

### **2.3. Превышение уровня шума**

Раздражающее влияние на работника оказывает шум, что является причиной повышения утомляемости, а при выполнении задач, требующих сосредоточенности и внимания, может привести к росту ошибок и увеличению длительности выполнения задания. Нормативным эквивалентным уровнем звука на рабочих местах (за исключением рабочих мест отдельных отраслей экономики) является 80 дБА [18].

В компьютерной аудитории № 418 КЦ ТПУ основными источниками шумов являются составляющие ПК. К таким составляющим относятся:

- вентилятор блока питания;
- вентилятор кулера центрального процессора;
- вентилятор на высокопроизводительной видеокарте;
- дополнительный вентилятор в корпусе системного блока.

Для понижения уровня шума необходимо понизить температуру внутри системного блока, что достигается достаточной вентиляцией системного блока. Для охлаждения системного блока необходимо

оборудовать со стороны вентиляционных отверстий хотя бы 20-30 см свободного пространства. К тому же необходимо не загромождать оборудование посторонними предметами, которые снижают теплоотдачу, а также прочищать вентиляционные отверстия от пыли.

С помощью мобильного приложения «Sound Meter» были произведены измерения уровня шума в помещении. Результаты измерений представлены в таблице 13.

Таблица 13 – Результаты измерений уровня звука

Рабочее место	Рабочая операция	Уровень звука, дБА		Продолжительность операции, мин	
		Результаты измерений	Эквивалентный уровень за операцию	Результаты наблюдений	Средняя
Аудитория КЦ ТПУ	Работа за персональным компьютером	43, 48, 45	45,3	20, 20, 20	20

В результате вычисления на основе измеренных величин показателей шума на рабочих местах эквивалентные уровни звука за 8-часовой рабочий день составляют 45,3 и 37,3 дБА, что соответствует нормативным значениям.

Снизить уровень шума в помещении можно с помощью звукопоглощающих материалов с максимальными коэффициентами звукопоглощения в области частот 63 – 8000 Гц для отделки стен и потолка помещений. Также дополнительный звукопоглощающий эффект создается однотонными занавесками из плотной ткани, повешенными в складку на расстоянии 15 – 20 см от ограждения. Ширина занавески должна быть в 2 раза больше ширины окна.

#### **2.4. Отсутствие или недостаток естественного и искусственного освещения**

В помещении при работе с ПК должно быть естественное и искусственное освещение. Естественное освещение обеспечивается через оконные проемы с коэффициентом естественного освещения не ниже 1,2% в зонах с устойчивым снежным покровом и не ниже 1,5% на остальной территории. Световой поток из оконного проема должен падать на рабочее

место оператора с левой стороны.

Для определения приемлемого уровня освещенности в помещении необходимо следующее: определить требуемый уровень освещенности внешними источниками света; если требуемый уровень освещенности не приемлем, необходимо найти способ сохранения требуемого контраста изображения другими средствами.

Рекомендуемые соотношения яркостей в поле зрения между рабочими поверхностями не должно превышать 1:3 – 1:5, между рабочими поверхностями и поверхностями стен и оборудования – 1:10.

Освещенность на поверхности стола в зоне размещения документа должна быть 300 – 500 лк [15]. Допускается установка светильников местного освещения для подсветки документов. Местное освещение не должно создавать бликов на поверхности экрана и увеличивать освещенность экрана более 300 лк. Прямую блеклость от источников освещения следует ограничить. Яркость светящихся поверхностей (окна, светильники), находящихся в поле зрения, должна быть не более 200 кд/м<sup>2</sup>. В таблице 14 представлены нормы на освещение для операторов поста управления [19].

Таблица 14 – Нормы на освещение для оператора

Характер зрительной работы	Разряд зрительной работы	Подразряд зрительной работы	Искусственное освещение		Естественное освещение, % при боковом направлении
			Освещенность при системе общего освещения, лк	Коэффициент пульсации, %	
Различие объектов высокой точности	Б	1	300	15	1

Для выявления соответствия компьютерной аудитории № 418 КЦ ТПУ требованиям СНиП 23-05-95\* далее будут приведены расчеты искусственного освещения рабочей зоны работника согласно методическим указаниям к

выполнению лабораторной работы «Расчет искусственного освещения» для студентов дневного и заочного обучения всех направлений и специальностей ТПУ [28].

Параметры компьютерной аудитории № 418 КЦ ТПУ:

- площадь  $S = 32,7 \text{ м}^2$ ;
- длина  $A = 6 \text{ м}$ ;
- ширина  $B = 5,45 \text{ м}$ ;
- высота потолка  $H = 2,5 \text{ м}$ ;
- высота рабочей поверхности  $h_{rp} = 0,8 \text{ м}$ ;
- количество светильников  $n = 12$ ;
- тип светильников ЛВО 4×18 CSVТ люминесцентными лампами типа L18W/640 с потоком  $F = 1200 \text{ лм}$ .

Согласно отраслевым нормам освещенности уровень рабочей поверхности над полом составляет 0,8 м., а установлена минимальная норма освещенности  $E = 300 \text{ лк}$  [29].

Расчетная высота подвеса светильников над рабочей поверхностью ( $h$ ) определяется по формуле (22):

$$h = H - h_p - h_c, \quad (22)$$

где:

$H$  – высота потолка в помещении, м;

$h_p$  – расстояние от пола до рабочей поверхности стола, м;

$h_c$  – расстояние от потолка до светильника, м.

Так как светильники установлены в уровень с потолком, примем  $h_c$  равным таким образом, согласно формуле (7.1) высота подвеса светильников над рабочей поверхностью для компьютерной аудитории № 418 КЦ ТПУ равна:

$$h = 2,5 - 0,8 = 1,7 \text{ м}.$$

Индекс помещения определяется по формуле (23):

$$i = \frac{S}{h(A+B)} \quad (23)$$

где:

$S$  – площадь помещения, м<sup>2</sup>;

$A$  – длина комнаты, м;

$B$  – ширина комнаты, м;

$h$  – высота подвеса светильников, м.

Индекс помещения для компьютерной аудитории № 418 КЦ ТПУ:

$$i = \frac{32,7}{1,7(6 + 5,45)} = 1,68$$

Исходя из того, что потолок в аудитории представлен светлыми потолочными панелями Armstrong, поверхность стен окрашена белой краской имеется два окна без штор, примем коэффициенты отражения от стен  $\rho_c = 50\%$  и потолка  $\rho_n = 50\%$ .

По таблице коэффициентов использования светового потока из пособия к МГСН 2.06-99 для соответствующих значений  $i$ ,  $\rho_c$ ,  $\rho_n$  примем  $\eta = 0,36$  (для потолочного светильника любого типа при  $i = 1,75$ ) [30].

Освещенность помещения рассчитывается по формуле (24):

$$E_\Phi = \frac{n \cdot \eta \cdot \Phi}{S \cdot k_z \cdot z}, \quad (24)$$

где

$\Phi$  – световой поток светильника, лм;

$S$  – площадь помещения, м<sup>2</sup>;

$k_z$  – коэффициент неравномерности освещения;

$n$  – число светильников;

$\eta$  – коэффициент использования светового потока.

Коэффициент запаса  $k$  учитывает запыленность светильников и их износ. Для помещений с малым выделением пыли  $k = 1,5$ . Поправочный коэффициент  $z$  – это коэффициент неравномерности освещения. Для люминесцентных ламп  $z = 1,1$ . Каждый светильник состоит из 4 ламп.

Учитывая все параметры, освещенность аудитории № 418 равна (24):

$$E_{\phi} = \frac{12 \cdot 4 \cdot 0,36 \cdot 1200}{32,7 \cdot 1,5 \cdot 1,1} = 384 \text{ лк}$$

Освещенность компьютерной аудитории №418 КЦ ТПУ удовлетворяет требованиям естественного и искусственного освещения [19].

## 2.5. Повышенная напряженность магнитного поля

Электромагнитные поля, характеризующиеся напряженностями электрических и магнитных полей, наиболее вредны для организма человека. Основным источником этих проблем, связанных с охраной здоровья людей, использующих в своей работе автоматизированные информационные системы на основе персональных компьютеров, являются дисплеи (мониторы), они представляют собой источники наиболее вредных излучений, неблагоприятно влияющих на здоровье человека. Предельно допустимые значения излучений от ЭВМ [15] приведены в таблице 15.

Таблица 15 – Допустимые уровни электромагнитного поля

Наименование параметров		ВДУ ЭМП
Напряженность электрического поля	В диапазоне частот 5 Гц – 2 кГц	25 В/м
	В диапазоне частот 2 кГц – 400 кГц	2,5 В/м
Плотность магнитного потока	В диапазоне частот 5 Гц – 2 кГц	250 нТл
	В диапазоне частот 2 кГц – 400 кГц	25 нТл
Электростатический потенциал экрана видеомонитора		500 В

Для оценки соблюдения уровней электромагнитного поля необходим производственный контроль. В случае их превышения необходимо проводить организационно-технические мероприятия (защита временем, расстоянием, экранирование источника, замена оборудования).

Для уменьшения уровня электромагнитного поля от персонального компьютера рекомендуется включать в одну розетку не более двух компьютеров, сделать защитное заземление, подключать компьютер к розетке через нейтрализатор электрического поля.

Помимо этого, для обеспечения более низкого уровня электромагнитного излучения на рабочем месте следует использовать жидкокристаллический монитор. Также в период работы с компьютером по возможности предусматривается сокращение времени, затрачиваемого на разработку и тестирование разрабатываемого ПО.

## 2.6. Экологическая безопасность

Деятельность по разработке ПО не связана с производством, поэтому

реализация поставленной задачи не влечет за собой негативных воздействий на окружающую среду. Это приводит к тому, что создание санитарно-защитной зоны и принятие мер по защите атмосферы, гидросферы, литосферы не являются необходимыми, а при работе над проектом применяются общие рекомендации по минимизации влияния на окружающую среду.

Единственным весомым исключением является использование ПК, чья неправильная утилизация может привести к загрязнению почвы или выбросам в атмосферу загрязняющих веществ, углекислого газа и выделение тепла в случае пожара. В конце срока службы ПК его можно отнести к отходам электронной промышленности. Переработка таких отходов осуществляется путем разделения на однородные компоненты, пригодных для дальнейшего использования, и их передачи для дальнейшего использования (например, кремния, алюминия, золота, серебра, редких металлов) [21].

Перечень элементов и содержащее их отработанное электротехническое и электронное оборудование, которые должны быть отдельно собраны при выводе отработавшего электротехнического и электронного оборудования из эксплуатации:

- конденсаторы, содержащие ПХБ;
- печатные платы и других устройств с площадью поверхности больше 10 см<sup>2</sup>;
- картриджи;
- пластик;
- электронно-лучевые трубки;
- элементы отработавшего электротехнического и электронного оборудования;
- газоразрядные лампы;
- жидкокристаллические экраны (если необходимо, вместе с корпусом) с поверхностью более 100 см<sup>2</sup> и все экраны с подсветкой газоразрядными лампами;



- внешние электрические кабели;
- элементы, содержащие огнеупорные керамические слои;
- конденсаторы, содержащие электролит (размер хотя бы одной из сторон конденсатора должен быть 25 мм или более).

Люминесцентные лампы относят к ртутьсодержащим отходам, и для их утилизации действует Постановление Правительства РФ [22]. Устанавливается порядок обращения с отходами производства и потребления в части осветительных устройств, электрических ламп, ненадлежащие сбор, накопление, использование, обезвреживание, транспортирование и размещение которых может повлечь причинение вреда жизни, здоровью граждан, вреда животным, растениям и окружающей среде.

Не допускается самостоятельное обезвреживание, использование, транспортирование и размещение отработанных ртутьсодержащих ламп потребителями отработанных ртутьсодержащих ламп, а также их накопление в местах, являющихся общим имуществом собственников помещений многоквартирного дома, за исключением размещения в местах первичного сбора и размещения, и транспортирования до них.

Сбор отработанных ртутьсодержащих ламп у потребителей осуществляют специализированные организации. Отходы, не подлежащие переработке и вторичному использованию, подлежат захоронению на полигонах.

### **3. Безопасность в чрезвычайных ситуациях**

#### **3.1. Анализ вероятных ЧС, которые может инициировать объект исследований**

Наиболее характерной ЧС для помещения, оборудованных ЭВМ, является пожар. Пожары на производстве возникают по определенным причинам, устранение которых составляет основу всех мероприятий по пожарной безопасности. Основные причины возникновения пожара [32]:

- нарушение правил эксплуатации электрического оборудования, эксплуатация его в неисправном состоянии;

- применение неисправных осветительных приборов, электропроводки и устройств, дающих искрение, замыкание и т. п.;
- перегрузка электрических сетей;

Аудитория КЦ ТПУ, относится к категории ВЗ по пожароопасности, содержит вещества и материалы, способные гореть при взаимодействии с водой, кислородом воздуха или друг с другом [23].

Для минимизации возможности возникновения пожара необходимо проводить пожарную профилактику. Под пожарной профилактикой понимают комплекс организационных и технических мероприятий, направленных на обеспечение безопасности людей, на предотвращении пожара, ограничение его распространения, а также создание условий для успешного тушения пожара. Для профилактики пожара чрезвычайно важна правильная оценка пожароопасности, определение опасных факторов и обоснование способов и средств предупреждения пожара и защиты от него.

Одно из условий обеспечения пожаробезопасности - ликвидация возможных источников воспламенения. Обогревание помещения открытыми

электронагревательными приборами могут привести к пожару, т.к. в помещении находятся бумажные документы и справочная литература. Следовательно, использование открытого нагревательного прибора неприемлемо.

### **3.2. Меры по предотвращению и ликвидации ЧС и их последствий**

В целях предотвращения пожара предлагаются следующие мероприятия:

- проведение с сотрудниками противопожарного инструктажа;
- проведение планового осмотра и своевременное устранение всех неисправностей в электроприборах;
- оснащение помещения автоматической системой обнаружения пожара;
- оснащение помещения автоматической системой оповещения о

пожаре.

Помещения с категорией ВЗ должны оснащаться системой автоматической пожарной сигнализации и система оповещения и управления эвакуацией людей при пожарах [24, 25].

Аудитория КЦ ТПУ оснащена одним пожарным, а также двумя дымовыми оповещателями, для информирования о пожаре установлена звуковая сирена.

### **Заключение по разделу**

Работник должен проводить профессиональную деятельность с учетом социальных, правовых, экологических и культурных аспектов, вопросов здоровья и безопасности, нести социальную ответственность за принимаемые решения, осознавать необходимость устойчивого развития.

В данном разделе были рассмотрены основные вопросы соблюдения прав работника на труд, выполнения правил к безопасности труда, промышленной безопасности, экологии и ресурсосбережения.

Установлено, что рабочее место исследователя удовлетворяет требованиям безопасности и гигиены труда во время реализации проекта, а вредное воздействие объекта исследования на окружающую среду не превышает норму.

## **ЗАКЛЮЧЕНИЕ**

В результате выполнения выпускной квалификационной работы были проанализированы существующие алгоритмы машинного обучения и разработаны ансамбли алгоритмов для обнаружения аномалий в медицинских исследованиях. Алгоритм создан для упрощения и ускорения работы медицинских работников, а также точности определения аномалий для постановки верного диагноза.

Проведен сравнительный анализ точности работы алгоритмов и ансамблей алгоритмов. В результате работы лучшим алгоритмом в рамках данной работы является ансамбль трёх алгоритмов «Elliptic Envelope + One Class SVM + Isolation Forest».

Научной новизной данной работы можно назвать использование ансамблей алгоритмов в области медицины для работы с большим объемом данных.

## **СПИСОК ПУБЛИКАЦИЙ И НАУЧНЫХ ДОСТИЖЕНИЙ**

1. Popenko Anastasia ANOMALIES IN THE RESULTS OF MEDICAL // RESEARCH VIII International Multidisciplinary Conference «Prospects and Key Tendencies of Science in Contemporary World» // Международная междисциплинарная конференция «Инновации и тенденции развития современной науки» // Bubok Publishing S.L., Madrid, Spain. 2021.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Anomaly Detection Principles and Algorithms 2017 Edition
2. Hands-On Machine Learning with scikit-learn and Scientific Python Toolkits (Released 7/24/2020)
3. Beginning Anomaly Detection Using Python-Based Deep Learning: With Keras and PyTorch 1st ed. 2019
4. Outlier Analysis 2nd ed. 2017 Edition
5. Outlier Detection: Techniques and Applications 1st Ed. 2019 Edition
6. Топ-5 языков для машинного обучения [Электронный ресурс] / С. Стельмах. — Электрон. текстовые дан. — 2019. — URL: <https://www.itweek.ru/ai/article/detail.php?ID=207784> (дата обращения 26.05.2021)
7. Kaggle [Электронный ресурс] URL:<https://ru.wikipedia.org/wiki/Kaggle> (дата обращения 25.05.2021)
8. Scikit-learn.org [Электронный ресурс] URL: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html>(дата обращения 20.03.2021)
9. Scikit-learn.org [Электронный ресурс] URL:[https://scikit-learn.org/stable/modules/generated/sklearn.svm.OneClassSVM.html?highlight=svm#sklearn.svm.OneClassSVM.decision\\_function](https://scikit-learn.org/stable/modules/generated/sklearn.svm.OneClassSVM.html?highlight=svm#sklearn.svm.OneClassSVM.decision_function)(дата обращения 10.04.2021)
10. Scikit-learn.org [Электронный ресурс] URL:<https://scikit-learn.org/stable/modules/generated/sklearn.covariance.EllipticEnvelope.html#sklearn.covariance.EllipticEnvelope> (дата обращения 16.04.2021)
11. История развития ансамблевых методов классификации в машинном обучении (работа Ю. Кашницкого) 2015г.
12. Ensemble methods: bagging, boosting and stacking [ Электронный ресурс ] URL: <https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205> (дата обращения 11.05.2021)

13. Трудовой кодекс Российской Федерации от 30.12.2001 N 197-ФЗ(ред. от 27.12.2018).
14. Федеральный закон от 28 декабря 2013 г. N 426-ФЗ «О специальной оценке условий труда».
15. СанПиН 2.2.2/2.4.1340–03. Санитарно-эпидемиологические правила и нормативы «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы» (с изменениями на 21 июня 2016 года).
16. ГОСТ 12.0.003-2015 ССБТ. Опасные и вредные производственные факторы. Классификация.
17. СанПиН 2.2.4.548–96. Гигиенические требования к микроклимату производственных помещений.
18. СанПиН 2.2.4.3359–16. Санитарно-эпидемиологические требования к физическим факторам на рабочих местах.
19. СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95.
20. ГОСТ Р 55090-2012 Ресурсосбережение. Обращение с отходами. Рекомендации по утилизации отходов бумаги.
21. ГОСТ Р 55102-2012 Ресурсосбережение. Обращение с отходами. Руководство по безопасному сбору, хранению, транспортированию и разборке отработавшего электротехнического и электронного оборудования, за исключением ртутьсодержащих устройств и приборов.
22. Постановление Правительства РФ от 03.09.2010 № 681 (ред. от 01.10.2013) Об утверждении Правил обращения с отходами производства и потребления в части осветительных устройств, электрических ламп, ненадлежащие сбор, накопление, использование, обезвреживание, транспортирование и размещение которых может повлечь причинение вреда жизни, здоровью граждан, вреда животным, растениям и окружающей среде.
23. НПБ 105-03. Определение категорий помещений, зданий и наружных

- установок по взрывопожарной и пожарной опасности.
24. СП 5.13130.2009. Системы противопожарной защиты. Установки пожарной сигнализации и пожаротушения автоматические. Нормы и правила проектирования (с Изменением № 1).
  25. СП 3.13130.2009. Системы противопожарной защиты. Система оповещения и управления эвакуацией людей при пожаре. Требования пожарной безопасности.
  26. ГОСТ 12.1.030–81 ССБТ Защитное заземление, зануление.
  27. ГОСТ Р 12.1.019-2009 ССБТ. Электробезопасность. Общие требования и номенклатура видов защиты.
  28. Безопасность жизнедеятельности. Расчёт искусственного освещения. Методические указания к выполнению индивидуальных заданий для студентов дневного и заочного обучения всех направлений и специальностей ТПУ. – Томск: Изд. ТПУ, 2008. – 20 с.
  29. СанПиН 2.2.2.542-96 Гигиенические требования к видео дисплейным терминалам, персональным электронно-вычислительным машинам и организации работ.
  30. Пособие к МГСН 2.06-99 Расчет и проектирование искусственного освещения помещений общественных зданий.
  31. Валуев Д.В., Гизатулин Р.А. Технологии переработки металлургических отходов: учебное пособие. – Томск: Издательство Томского политехнического университета, 2012. – 196 с.
  32. Лабораторный практикум по дисциплине «Безопасность жизнедеятельности» для студентов всех специальностей: учебное пособие. Ю.А. Амелькович, Ю.В. Анищенко, А.Н. Вторушина, М.В. Гуляев, М.Э. Гусельников, А.Г. Дашковский, Т.А. Задорожная, В.Н. Извеков, А.Г. Кагиров, К.М. Костырев, В.Ф. Панин, А.М. Плахов, С.В. Романенко – Томск: Издательство Томского политехнического университета, 2010. – 236 с



33. ГОСТ Р 22.0.01-2016 Безопасность в чрезвычайных ситуаци

## ПРИЛОЖЕНИЕ А

### Data preparation and economic indicators calculation

Студент:

Группа	ФИО	Подпись	Дата
8ПМ9И	Попенко Анастасия Александровна		

Руководитель ИШИТР ОИТ:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Аксёнов Сергей Владимирович	К.Т.Н.		

Консультант – лингвист ШБИП ОИЯ:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИЯ	Сидоренко Татьяна Валерьевна	К.П.Н.		

## **Object and research methods**

### **1.1. Machine learning techniques**

The first methods came from pure statistics in the '50s. They solved formal math tasks — searching for patterns in numbers, evaluating the proximity of data points, and calculating vectors' directions.

The only goal of machine learning is to predict results based on incoming data. The greater variety in the samples you have, the easier it is to find relevant patterns and predict the result. Therefore, we need three components to teach the machine (Figure 1):

#### **1.Data**

There are two main ways to get the data — manual and automatic. Manually collected data contains far fewer errors but takes more time to collect — that makes it more expensive in general. Automatic approach is cheaper — you're gathering everything you can find and hope for the best.

#### **2. Features**

Also known as parameters or variables. Those could be car mileage, user's gender, stock price, word frequency in the text. In other words, these are the factors for a machine to look at.

#### **3. Algorithms**

Most obvious part. Any problem can be solved differently. The method you choose affects the precision, performance, and size of the final model. There is one important nuance though: if the data is crappy, even the best algorithm won't help. Sometimes it's referred as "garbage in – garbage out". So don't pay too much attention to the percentage of accuracy, try to acquire more data first.

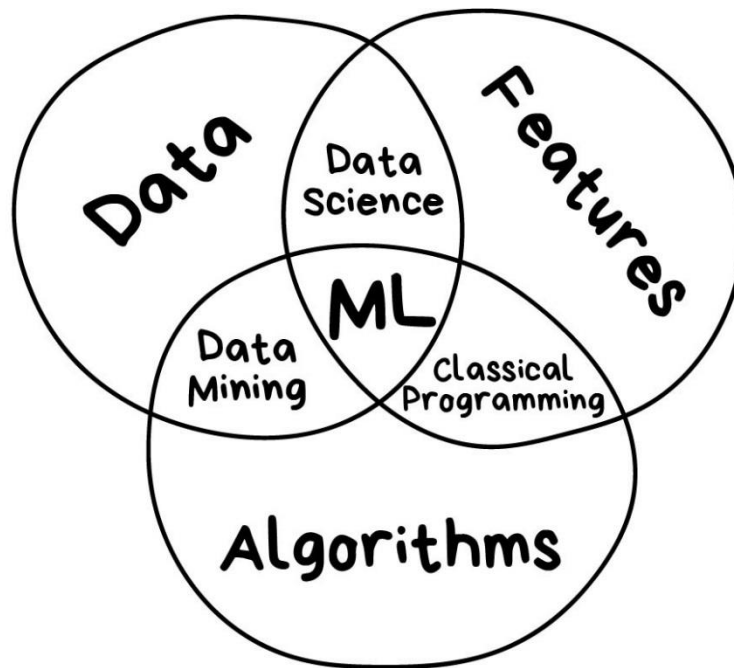


Figure 1 Components to teach the machine

### 1.1.1. Supervised Learning

Today used for:

- Spam filtering
- Language detection
- Fraud detection
- Stock price forecasts
- Medical diagnosis
- Any number-time correlations

How it works: This algorithm consist of a target / outcome variable (or dependent variable) which is to be predicted from a given set of predictors (independent variables). Using these set of variables, we generate a function that map inputs to desired outputs. The training process continues until the model achieves a desired level of accuracy on the training data. Examples of Supervised Learning: Regression, Decision Tree, Random Forest, KNN, Logistic Regression etc.

### 1.1.2. Unsupervised Learning

Nowadays used:

- To analyze and label new data
- To detect abnormal behavior

- To analyze web surfing patterns
- Topic modeling and similar document search
- Fake image analysis
- Risk management

How it works: In this algorithm, we do not have any target or outcome variable to predict / estimate. It is used for clustering population in different groups, which is widely used for segmenting customers in different groups for specific intervention. Examples of Unsupervised Learning: Apriori algorithm, K-means.

### **1.1.3. Reinforcement Learning**

Nowadays used for:

- Self-driving cars
- Robot vacuums
- Games
- Automating trading
- Enterprise resource management

How it works: Using this algorithm, the machine is trained to make specific decisions. It works this way: the machine is exposed to an environment where it trains itself continually using trial and error. This machine learns from past experience and tries to capture the best possible knowledge to make accurate business decisions. Example of Reinforcement Learning: Markov Decision Process.

### **1.1.4. Deep Learning**

Used today for:

- Replacement of all algorithms above
- Object identification on photos and videos
- Speech recognition and synthesis
- Image processing, style transfer
- Machine translation

Deep Learning is a modern method of building, training, and using neural networks. Basically, it's a new architecture. Nowadays in practice, no one separates deep learning from the "ordinary networks". We even use the same libraries for them. The general rule is to compare things on the same level.

### **1.1.5. Reinforcement Learning**

Nowadays used for:

- Self-driving cars
- Robot vacuums
- Games
- Automating trading
- Enterprise resource management

Reinforcement learning is used in cases when your problem is not related to data at all, but you have an environment to live in. Like a video game world or a city for self-driving car.

Knowledge of all the road rules in the world will not teach the autopilot how to drive on the roads. Regardless of how much data we collect, we still can't foresee all the possible situations. This is why its goal is to minimize error, not to predict all the moves. Surviving in an environment is a core idea of reinforcement learning.

### **1.1.6. Ensemble Methods**

Nowadays is used for:

- Everything that fits classical algorithm approaches (but works better)
- Search systems
- Computer vision
- Object detection

Ensembles and neural networks are two main fighters paving our path to a singularity. Today they are producing the most accurate results and are widely used in production.

Despite all the effectiveness the idea behind these is overly simple. If you take a bunch of inefficient algorithms and force them to correct each other's mistakes, the overall quality of a system will be higher than even the best individual algorithms.

You'll get even better results if you take the most unstable algorithms that are predicting completely different results on small noise in input data. Like

Regression and Decision Trees. These algorithms are so sensitive to even a single outlier in input data to have models go mad.

A visualization of the machine learning methods described above is shown in Figure 2.

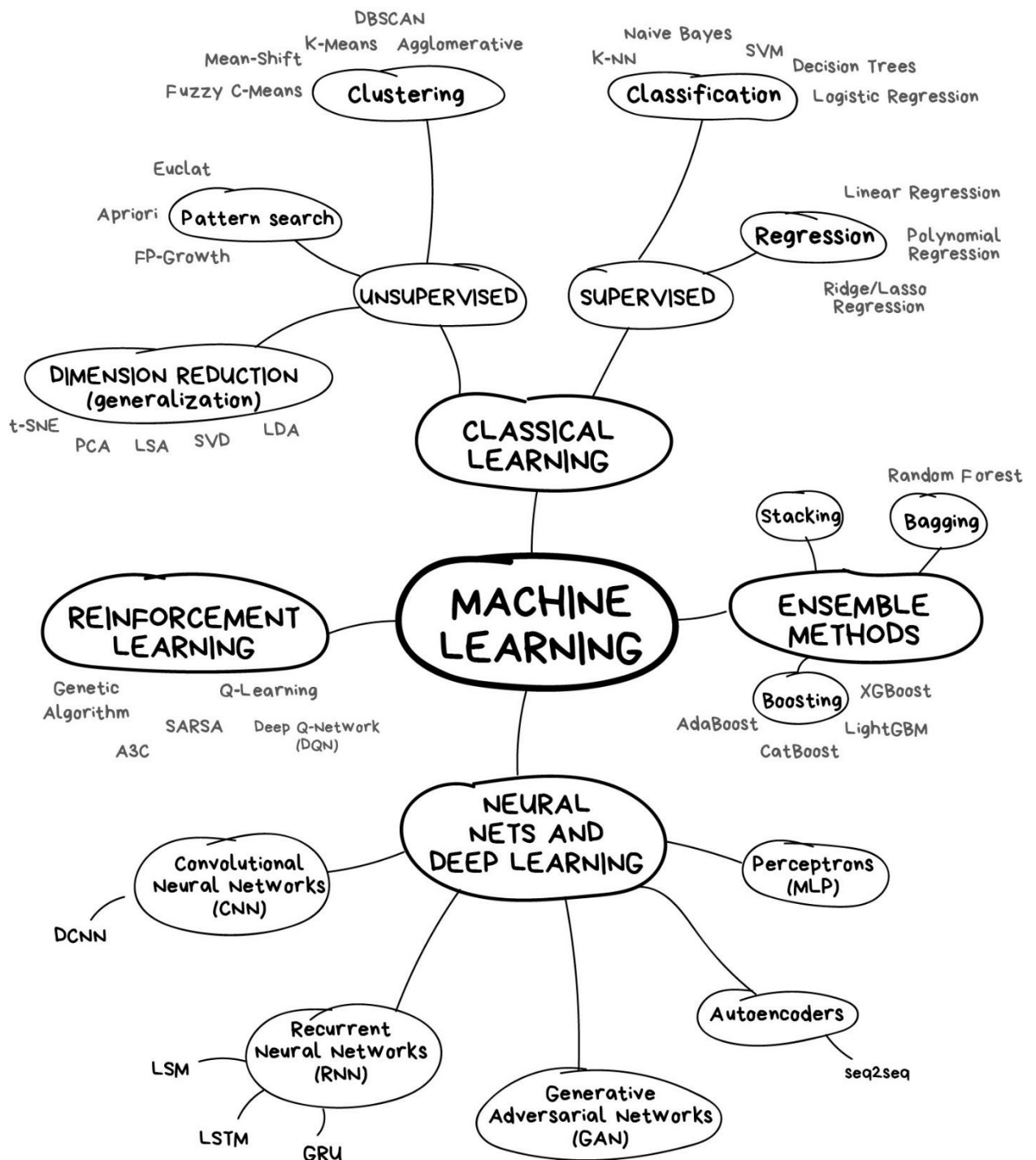


Figure 2 Machine learning methods

## ПРИЛОЖЕНИЕ Б

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn import svm
from sklearn.covariance import EllipticEnvelope
import seaborn as sns
from matplotlib import rc
df0=pd.read_csv('all_weekly_excess_deaths.csv')
df=df0[['week','covid_deaths']]
x=df.values
from sklearn.ensemble import IsolationForest
forest=IsolationForest(random_state=0)
forest.fit(x)
y = forest.predict(x)
print(y)
outliers_fraction = 0.25
svmloutl=svm.OneClassSVM(nu=0.95 * outliers_fraction +
0.05, kernel="rbf", gamma=0.1)
svmloutl.fit(x)
svmy = svmloutl.predict(x)
print(svmy)
el=EllipticEnvelope(contamination=outliers_fraction)
el.fit(x)
ely = el.predict(x)
print(ely)
from sklearn.cluster import DBSCAN
dbs = DBSCAN(eps=0.3, min_samples=10).fit(x)
df['outlierForest']=y
df['outliersvm']=svmy
df['outlierel']=ely
```



```

df.to_excel('out.xlsx')
midwest=pd.read_excel('out.xlsx')

sns.scatterplot(x="week", y="covid_deaths", hue="outlierForest",
data=df)

sns.scatterplot(x="week", y="covid_deaths", hue="outlierel",
data=df)

df['pic1']=df['outliersvm']+df['outlierel']
df.to_excel('out22.xlsx')
pic=pd.read_excel('out22.xlsx')
df['2 метода SVM+EE']=''
mask1=df['pic1']==-2
mask2=df['pic1']==0
mask3=df['pic1']==2
df['2 метода SVM+EE'][mask1]='выброс2'
df['2 метода SVM+EE'][mask2]='выброс1'
df['2 метода SVM+EE'][mask3]='выборка'
sns.scatterplot(data=df, x="week", y="covid_deaths", hue="2
метода SVM+EE")

df['pic10']=df['outliersvm']+df['outlierForest']
df.to_excel('out44.xlsx')
pic=pd.read_excel('out44.xlsx')
df['2 метода SVM+IF']=''
mask1=df['pic10']==-2
mask2=df['pic10']==0
mask3=df['pic10']==2
df['2 метода SVM+IF'][mask1]='выброс2'
df['2 метода SVM+IF'][mask2]='выброс1'
df['2 метода SVM+IF'][mask3]='выборка'
sns.scatterplot(data=df, x="week", y="covid_deaths", hue="2
метода SVM+IF")

df['pic110']=df['outlierel']+df['outlierForest']
df.to_excel('out410.xlsx')
pic=pd.read_excel('out410.xlsx')

```

```

df['2 метода EE+IF']=''
mask1=df['pic110']==-2
mask2=df['pic110']==0
mask3=df['pic110']==2
df['2 метода EE+IF'][mask1]='выброс1'
df['2 метода EE+IF'][mask2]='выброс2'
df['2 метода EE+IF'][mask3]='выборка'
sns.scatterplot(data=df, x="week", y="covid_deaths", hue="2
метода EE+IF")
df['pic100']=df['outlierel']+df['outlierForest']+df['outliersvm'
]
df.to_excel('out400.xlsx')
pic=pd.read_excel('out400.xlsx')
df['3 метода']=''
mask1=df['pic100']==-3
mask2=df['pic100']==-1
mask3=df['pic100']==3
mask4=df['pic100']==1
df['3 метода'][mask1]='выброс3'
df['3 метода'][mask2]='выброс2'
df['3 метода'][mask3]='выборка'
df['3 метода'][mask4]='выброс1'
sns.scatterplot(data=df, x="week", y="covid_deaths", hue="3
метода")
# Confusion Matrix
from sklearn.metrics import confusion_matrix
confusion_matrix(y, svmy)
a=[[ 350, 349],[732, 1764]]
df=pd.DataFrame(a, range(2), range(2))
sns.heatmap(df)
plt.show()
from sklearn.metrics import accuracy_score

```

```

accuracy_score(ely, y)
confusion_matrix(ely, svmy)
b=[[ 414, 385],[ 668, 1728]]
df=pd.DataFrame(a, range(2), range(2))
sns.heatmap(df)
plt.show()
confusion_matrix(ely, y)
c=[[ 596, 203],[103, 2293]]
df2=pd.DataFrame(c, range(2), range(2))
sns.heatmap(df2)
plt.show()
# Accuracy
from sklearn.metrics import accuracy_score
accuracy_score(ely, y)
# Recall
from sklearn.metrics import recall_score
recall_score(ely, y, average=None)
# Precision
from sklearn.metrics import precision_score
precision_score(ely, y, average=None)
#F1score
from sklearn.metrics import f1_score
f1_score(ely, y, average=None)
#Classification_report
from sklearn.metrics import classification_report
print(classification_report(ely, y))

```

## ПРИЛОЖЕНИЕ В



Director General  
Publishing House "Internauka"

  
I. Gulin

25.06.2021  
Madrid, Spain



Рисунок В1 Сертификат участника конференции

# ИНТЕРНАУКА

ИНН 7715986433, р/сч № 40702810929140005840 в ФИЛИАЛ "НИЖЕГОРОДСКИЙ" АО "АЛЬФА-БАНК"  
кор/сч 30101810200000000824 БИК 042202824  
E-mail: mail@internauka.org, сайт: www.internauka.org, тел.8(499)-938-55-90,  
адрес: 125424, Москва, Волоколамское шоссе, д. 108, помещение VIII, комн. 4, офис 33

## CONFIRMATION LETTER / СПРАВКА

21.06.2021 № 56273

Автор(-ы): Popenko Anastasia Alexsandrovna

The publishing house "Internauka" informs you that your article entitled "ANOMALIES IN THE RESULTS OF MEDICAL RESEARCH" has been accepted for publication in the proceedings of the VIII International Multidisciplinary Conference "Prospects and Key Tendencies of Science in Contemporary World". The metadata is to place on eLIBRARY.ru.

The publication on the website <https://www.internauka.org/> is due on July 02, 2021.

Издательство «Интернаука» сообщает Вам о том, что Ваша статья «ANOMALIES IN THE RESULTS OF MEDICAL RESEARCH» принята к публикации в сборнике статей VIII International Multidisciplinary Conference "Prospects and Key Tendencies of Science in Contemporary World". Метаданные сборника статей будут размещены на eLIBRARY.ru.

Планируемая дата публикации сборника статей конференции на сайте <https://www.internauka.org/> 02.07.2021 г.

Генеральный директор  
Издательство «Интернаука»

И.А. Гулин



Рисунок В2 Справка принятия статьи к публикации